8-31-2021

Novel statistical modeling methods for traffic video analysis

Hang Shi
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

NOVEL STATISTICAL MODELING METHODS FOR TRAFFIC VIDEO ANALYSIS

by
Hang Shi

Video analysis is an active and rapidly expanding research area in computer vision and artificial intelligence due to its broad applications in modern society. Many methods have been proposed to analyze the videos, but many challenging factors remain untackled. In this dissertation, four statistical modeling methods are proposed to address some challenging traffic video analysis problems under adverse illumination and weather conditions.

First, a new foreground detection method is presented to detect the foreground objects in videos. A novel Global Foreground Modeling (GFM) method, which estimates a global probability density function for the foreground and applies the Bayes decision rule for model selection, is proposed to model the foreground globally. A Local Background Modeling (LBM) method is applied by choosing the most significant Gaussian density in the Gaussian mixture model to model the background locally for each pixel. In addition, to mitigate the correlation effects of the Red, Green, and Blue (RGB) color space on the independence assumption among the color component images, some other color spaces are investigated for feature extraction. To further enhance the discriminatory power of the input feature vector, the horizontal and vertical Haar wavelet features and the temporal information are integrated into the color features to define a new 12-dimensional feature vector space. Finally, the Bayes classifier is applied for the classification of the foreground and the background pixels.

Second, a novel moving cast shadow detection method is presented to detect and remove the cast shadows from the foreground. Specifically, a set of new chromatic criteria is presented to detect the candidate shadow pixels in the Hue, Saturation, and Value (HSV) color space. A new shadow region detection method is then proposed to cluster the

candidate shadow pixels into shadow regions. A statistical shadow model, which uses a single Gaussian distribution to model the shadow class, is presented to classify shadow pixels. Additionally, an aggregated shadow detection strategy is presented to integrate the shadow detection results and remove the shadows from the foreground.

Third, a novel statistical modeling method is presented to solve the automated road recognition problem for the Region of Interest (RoI) detection in traffic video analysis. A temporal feature guided statistical modeling method is proposed for road modeling. Additionally, a model pruning strategy is applied to estimate the road model. Then, a new road region detection method is presented to detect the road regions in the video. The method applies discriminant functions to classify each pixel in the estimated background image into a road class or a non-road class, respectively. The proposed method provides an intra-cognitive communication mode between the RoI selection and video analysis systems.

Fourth, a novel anomalous driving detection method in videos, which can detect unsafe anomalous driving behaviors is introduced. A new Multiple Object Tracking (MOT) method is proposed to extract the velocities and trajectories of moving foreground objects in video. The new MOT method is a motion-based tracking method, which integrates the temporal and spatial features. Then, a novel Gaussian Local Velocity (GLV) modeling method is presented to model the normal moving behavior in traffic videos. The GLV model is built for every location in the video frame, and updated online. Finally, a discriminant function is proposed to detect anomalous driving behaviors.

To assess the feasibility of the proposed statistical modeling methods, several popular public video datasets, as well as the real traffic videos from the New Jersey Department of Transportation (NJDOT) are applied. The experimental results show the effectiveness and feasibility of the proposed methods.

NOVEL STATISTICAL MODELING METHODS FOR TRAFFIC VIDEO ANALYSIS

by
Hang Shi

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

August 2021

APPROVAL PAGE

NOVEL STATISTICAL MODELING METHODS FOR TRAFFIC VIDEO ANALYSIS

Hang Shi

Dr. Chengjun Liu, Dissertation Advisor — Date
Professor of Computer Science, NJIT

Dr. James Geller, Committee Member — Date
Professor of Computer Science, NJIT

Dr. Ali Mili, Committee Member — Date
Professor of Computer Science, NJIT

Dr. Zhi Wei, Committee Member — Date
Professor of Computer Science, NJIT

Dr. Taro Narahara, Committee Member — Date
Associate Professor of Architecture and Design, NJIT

BIOGRAPHICAL SKETCH

Author: Hang Shi

Degree: Doctor of Philosophy

Date: August 2021

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2021
- Master of Science in Electrical Engineering,
  New York University, Brooklyn, NY, 2014
- Bachelor of Science in Automation,
  University of Science and Technology of China, Anhui, China, 2012

Major: Computer Science

Presentations and Publications:

Hadi Ghahremannezhad, Hang Shi, and Chengjun Liu, A New Online Approach for Moving Cast Shadow Suppression in Traffic Videos. In *24th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, September 19-22, 2021, Indianapolis, IN, USA.

Hang Shi, Hadi Ghahremannezhad, and Chengjun Liu, Anomalous Driving Detection for Traffic Surveillance Video Analysis. In *IEEE International Conference on Imaging Systems and Techniques (IST)*, August 24-26, 2021, New York, NY, USA.

Hang Shi and Chengjun Liu, An Innovative Video Quality Assessment Method and An Impairment Video Dataset. In *IEEE International Conference on Imaging Systems and Techniques (IST)*, August 24-26, 2021, New York, NY, USA.

Hang Shi and Chengjun Liu, A New Cast Shadow Detection Method for Traffic Surveillance Video Analysis Using Color and Statistical Modeling. *Image and Vision Computing*, 94, 103863, 2020.

Hadi Ghahremannezhad, Hang Shi, and Chengjun Liu, Automatic Road Detection in Traffic Videos. In *10th IEEE International Conference on Big Data and Cloud Computing (BDCloud)*, December 17-19, 2020, virtual.

Hadi Ghahremannezhad, Hang Shi, and Chengjun Liu, A New Adaptive Bidirectional Region-of-Interest Detection Method for Intelligent Traffic Video Analysis. In *3rd IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, December 9-11, 2020, virtual.

Hadi Ghahremannezhad, Hang Shi, and Chengjun Liu, Robust Road Region Extraction in Video Under Various Illumination and Weather Conditions. In *4th IEEE International Conference on Image Processing, Applications and Systems (IPAS)*, December 9-11, 2020, Genova, Italy.

Hang Shi, Hadi Ghahremannezhad, and Chengjun Liu, A Statistical Modeling Method for Road Recognition in Traffic Video Analytics. In *11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, September 23-25, 2020, virtual.

Hadi Ghahremannezhad, Hang Shi, and Chengjun Liu, A Real Time Accident Detection Framework for Traffic Video Analysis. In *16th International Conference on Machine Learning and Data Mining (MLDM)*, July 18-23, 2020, virtual.

Hang Shi and Chengjun Liu, Moving Cast Shadow Detection in Video Based on New Chromatic Criteria and Statistical Modeling. In *18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, December 16-19, 2019, Boca Raton, FL, USA.

Hang Shi and Chengjun Liu, An Unsupervised Statistical Moving Shadow Detection Method for Video Analysis. In *15th International Conference on Machine Learning and Data Mining (MLDM)*, July 20-25, 2019, New York, NY, USA.

Abbas Kiani, Guanxiong Liu, Hang Shi, Abdallah Khreishah, Nirwan Ansari, Jo Young Lee, and Chengjun Liu, A Two-Tier Edge Computing Based Model for Advanced Traffic Detection. In *5th International Conference on Internet of Things: Systems, Management and Security (IoTSMS)*, October 15-18, 2018, Valencia, Spain.

Hang Shi and Chengjun Liu, A New Foreground Segmentation Method for Video Analysis in Different Color Spaces. In *24th International Conference on Pattern Recognition (ICPR)*, August 20-24, 2018, Beijing, China.

Hang Shi and Chengjun Liu, A New Global Foreground Modeling and Local Background Modeling Method for Video Analysis. In *14th International Conference on Machine Learning and Data Mining (MLDM)*, July 14-19, 2018, New York, NY, USA.

To My Beloved Parents Yuzhan Shi and Xiaochun Huang, Parents-in-law Zhinan Jiang and Xiuzhi Ma, and My Wife Wei Jiang.

ACKNOWLEDGMENT

Foremost, I would like to express my heartfelt appreciation to my dissertation advisor, Dr. Chengjun Liu, for his invaluable advice, technical guidance, patience, and kindness to bring this dissertation proposal to fruition. Over the past few years, Dr. Liu has been a constant source of encouragement and support which has helped me to present our research achievements in top conferences and reputed journals.

I am extremely grateful to Dr. James Geller, Dr. James Mchugh, Dr. Ali Mili, Dr. Zhi Wei, and Dr. Taro Narahara for serving on my committee. I want to thank them for the time they have spent providing me with their valuable suggestions on my research.

I would also like to thank all my fellow graduate students Ajit Puthenputhussery, Hao Liu, Mohammad Omar Faruque, and Hadi Ghahremannezhad for their support and assistance.

I must also thank Ms. Angel Butler and Dr. Reza Curtmola in the Computer Science department for providing me with valuable academic advice and helping me with different academic issues.

I am thankful to the RA support from the National Science Foundation and New Jersey Department of Transportation for supporting my research work.

Most importantly, I heartily appreciate my family for standing by my side for every decision that I have taken in my life. I thank my parents, Mr. Yuzhan Shi, and Mrs. Xiaochun Huang, for their faith in all my ambitions. I am thankful to my wife Wei Jiang for her affection and encouragement during the tough years of Ph.D. study. I would also like to thank my parents-in-law, Mr. Zhinan Jiang, and Mrs. Xiuzhi Ma, for their support and encouragement.

Finally, I would also like to thank all my friends for enriching my graduate life, outside research.

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1
2	BACKGROUND AND RELATED WORK	7
	2.1 Foreground Detection Methods	7
	2.2 Shadow Detection Methods	9
	2.3 Road Region Detection Methods	10
	2.4 Anomalous Driving Detection Methods	12
3	A NOVEL FOREGROUND DETECTION METHOD FOR VIDEO ANALYSIS BY INTEGRATING COLOR, WAVELET, AND TEMPORAL FEATURES	15
	3.1 Introduction	15
	3.2 A Novel Foreground Detection Method for Video Analysis	17
	3.2.1 A Local Background Modeling (LBM) Method Using a Single Gaussian Density	18
	3.2.2 A Novel Global Foreground Modeling (GFM) Method	19
	3.2.3 Foreground and Background Classification	21
	3.2.4 Feature Vectors in Various Color Spaces	25
	3.2.5 Enhancing Discriminatory Power of the Feature Vector by Integrating Color, Wavelet, and Temporal Features	28
	3.3 Experiments	30
	3.4 Conclusions	41
4	A NEW MOVING CAST SHADOW DETECTION METHOD FOR VIDEO ANALYSIS USING COLOR AND STATISTICAL MODELING	43
	4.1 Introduction	43
	4.2 A Novel Moving Cast Shadow Detection Method	44
	4.2.1 New Chromatic Criteria for Shadow Pixel Detection	47
	4.2.2 A New Shadow Region Detection Method	52
	4.2.3 A New Statistical Shadow Modeling and Classification Method	54

TABLE OF CONTENTS
(Continued)

Chapter Page

4.2.4 Aggregated Shadow Detection 56

4.3 Experiments 58

4.4 Conclusions 63

5 A STATISTICAL MODELING METHOD FOR ROAD RECOGNITION IN TRAFFIC VIDEO ANALYTICS 64

5.1 Introduction 64

5.2 A Temporal Feature Guided Statistical Modeling Method for Road Recognition 66

5.2.1 The New Road Model Estimation Method 66

5.2.2 The Novel Road Recognition Method 69

5.3 Experiments 70

5.4 Conclusions 73

6 AN INNOVATIVE ANOMALOUS DRIVING DETECTION METHOD IN VIDEO 75

6.1 Introduction 75

6.2 The New Multiple Object Tracking Method Based on Temporal and Spatial Features 77

6.3 The Anomalous Driving Detection using a Novel Gaussian Local Velocity Model 80

6.4 Experiments 81

6.5 Conclusion 84

7 CONCLUSION AND FUTURE WORK 86

REFERENCES 89

LIST OF TABLES

Table		Page
3.1	The Run Time of the Foreground Detection Method using the Three Types of the NJDOT Traffic Videos: 352×240 Video, 704×480 Video, and 752×480 Video	30
3.2	The Foreground Detection Performance of Our Proposed Foreground Detection Method using Different Feature Vectors	33
3.3	The Foreground Detection Performance of Different Methods using the CD.NET-2014 Videos	35
3.4	The Foreground Detection Performance of Different Methods	37
3.5	The Description of the Video Sequences We used in Experiments	41
4.1	The Comparative Shadow Detection Performance of Our Proposed Method and Some Popular Shadow Detection Methods	59
5.1	The Quantitative Performance of the Proposed Method	72

LIST OF FIGURES

Figure		Page
1.1	Some example frames from traffic videos showing low video quality.	2
1.2	Some example frames from different traffic videos showing strong cast shadows.	3
1.3	Some example frames showing that the road regions only occupy a small portion of the frame.	5
1.4	An example frame showing that a vehicle is driving in a anomalous direction.	6
3.1	A comparison between a single Gaussian density with multiple Gaussian densities and a uniform distribution in foreground detection. The single Gaussian density has a smaller overlapping with the background model, which means a lower error rate.	22
3.2	The weight of the background Gaussian density will decrease when an object stops moving. The background model will be replaced with a new Gaussian distribution which represents the stopped moving object.	23
3.3	In Hayman and Eklundh's method, after an object stops moving, the foreground model will no longer remain any information of that foreground object. . .	24
3.4	Color component images in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The first row shows the R, G, and B component images in the RGB color space, the second row displays the Y, I, and Q component images in the YIQ color space, the third row displays the Y, Cb, and Cr component images in the YCbCr color space, the fourth row shows the three uncorrelated component images in the UCS, the fifth row shows the three independent component images in the ICS, and the sixth row shows the three discriminating component images in the DCS.	27
3.5	The feature vectors in different color spaces. Each column shows the 12-dimensional features in one color space, namely, the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. Each column contains the three color components, the horizontal Haar wavelet features, the vertical Haar wavelet features, and the temporal difference features in three corresponding colors.	29

LIST OF FIGURES
(Continued)

Figure **Page**

3.6 Comparative foreground detection performance of the proposed foreground detection method in three-dimensional feature paces and 12-dimensional feature spaces. The First row shows one video frame from an NJDOT traffic video with spatial resolution of 352×240 and the ground truth of the foreground mask. The second and the third rows show the foreground detection results using the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The fourth and the fifth rows show the foreground detection results using the 12-dimensional feature vectors constructed from the corresponding color spaces, respectively. 31

3.7 Comparative performance of the proposed foreground detection method and other popular methods using the CD.net-2014 videos. Each row represents a different video. From top to bottom are highway, office, pedestrians, and PETS2006, respectively. The First column shows some video frames from the data set. The second column shows the ground truth. The third to the seventh columns display the detection results using the BMOG method [77], the CL-VID method [73], the PAWCS method [106], the SBBS method [114], and the SWCD method [51], respectively. The last column displays the detection results of our proposed method in the ICS. 34

3.8 Comparative foreground detection performance of the proposed foreground detection method and some popular video analysis methods. (a) One video frame from an NJDOT traffic video with a spatial resolution of 352×240. (b) The ground truth of the foreground mask. (c)-(h) The foreground masks that are extracted using the proposed method in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. (i)-(l) The foreground masks which are extracted using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106], respectively. . . . 36

3.9 Comparative performance of the proposed foreground detection method and other popular methods in a scene with some stopped vehicles. The First row shows one video frame from an NJDOT traffic video with spatial resolution of 704×480, and the detection result of our proposed method. The second row and the third row display the detection results using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106], respectively. 38

3.10 Some stopped vehicle detection results. The stopped vehicles detected by our proposed method are marked with the red rectangles. 39

4.1 The system architecture of our novel moving cast shadow detection method. . 45

LIST OF FIGURES
(Continued)

Figure **Page**

4.2 (a) A video frame from an NJDOT traffic video. (b) The background derived using the GMM model. (c) The foreground (with shadow) detected using our new foreground detection method. 46

4.3 The HSV color space can be modeled as a cone. 48

4.4 (a) The H (hue) component of the video frame. (b) The S (saturation) component of the video frame. (c) The V (value) component of the video frame. (d) The difference of the S component between the frame and the background. (e) The difference of the V component between the background and the frame. 49

4.5 (a) A video frame from an NJDOT traffic video. (b) The shadow detection results (shadow pixels are represented using gray scale value of 128) using the chromatic criteria in [94], [38]. (c) The shadow detection results using our new chromatic criteria. 51

4.6 The color of the outline is different from that of the main part of the shadow. . 52

4.7 (a) A video frame from an NJDOT traffic video. (b) The shadow detection results using Huang and Chen's method [49]. (c) The shadow detection results using our new chromatic criteria. (d) The shadow detection results using our shadow region detection method.. 53

4.8 (a) A video frame from the NJDOT traffic video. (b) The detected foreground (with shadow) using the new foreground detection method. (c) The detected shadow pixels using the new chromatic criteria. (d) The detected shadow regions using the shadow region detection method. (e) The detected shadow pixels using statistical shadow modeling and classification. (f) The detected shadow pixels using the aggregated shadow detection method. (g) The shadow free foreground. (h) The video frame that shows foreground pixels in red color. 57

4.9 The foreground masks obtained by different methods. (a). One video frame of 'Highway-3' video [94]. (b). The ground truth of the foreground mask. The white parts are the foreground objects. The gray parts are the cast shadows. (c)-(g) The shadow free foreground mask of Cucchiara *et al.*'s method [29], Huang and Chen's method [49], Hsieh *et al.*'s method [48], Leone and Distante's method [62], and Sanin *et al.*'s method [95], respectively. (h)The shadow free foreground mask of our proposed method. 60

4.10 The comparison of vehicle tracking performance using a frame from the NJDOT traffic videos. (a) The vehicle tracking results without shadow detection. (b) The vehicle tracking results with shadow detection using our proposed shadow detection method. 62

LIST OF FIGURES
(Continued)

Figure **Page**

5.1 (a) shows a video frame from an NJDOT traffic video. (b) displays a binary foreground mask, where white pixels represent the moving foreground objects. (c) shows the estimated background image. (d) shows the corresponding regions of the moving foreground objects projected on the background image in red color. 67

5.2 The road recognition results. The first and the fifth rows display one video frame from an NJDOT traffic video. The second and the sixth rows show the ground truth road regions. The third and the seventh rows present the road recognition result of UFL-HS method. The fourth and the eighth rows show the road region detected by our proposed method. 71

5.3 The F-measure score of our proposed road recognition method using different number of training frames. 73

6.1 The workflow of our proposed anomalous driving detection method using the new MOT method and the novel GLV model. 76

6.2 Two adjacent video frames from the NJDOT video and their corresponding foreground masks. 78

6.3 The moving trajectories detected by our new MOT method. The red lines indicate the moving trajectories of the vehicles in two seconds. 82

6.4 The wrong-way driving vehicles detected by our proposed method. The red rectangles are the wrong-way driving vehicles detected by our proposed method, the areas in the green lines are the region of interest (ROI). 83

CHAPTER 1

INTRODUCTION

Video analysis, which has broad applications in surveillance and security, often applies popular techniques in multidisciplinary fields, such as computer vision, pattern recognition, machine learning, and artificial intelligence [14], [103]. Many statistical methods are proposed to analyze the video data, but there are still many challenging problems that need to be further investigated. Foreground detection is one of the most extensively studied topics in video analysis. One major problem in foreground detection is that the quality of some videos is fairly low. Many object detection methods, such as YOLO and mask RCNN can not detect the foreground objects due to the low resolution and poor quality. Figure. 1.1 shows some example frames from different traffic videos with low video quality. We can see that the quality of the videos is poor and the vehicles in these videos are hard to be recognized, which is the reason why most of the foreground detection methods can not reach high detection accuracy. The second issue of current foreground detection methods is the detection of the stopped moving foreground objects. Most foreground detection methods cannot keep detecting the foreground objects when they stop moving, or can only detect those objects for a short period of time. However, some real world applications are interested in the stopped moving foreground objects, and need foreground detection methods to detect them, such as the industrial production line monitoring system, and the traffic incident detection system.

Another challenging issue in traffic video analysis is the cast shadow detection problem. The cast shadows are often detected as part of the foreground since the cast shadows share similar motion patterns to the foreground objects [89], [94], [75]. Figure. 1.2 shows some example frames from different traffic videos with strong cast shadows. These

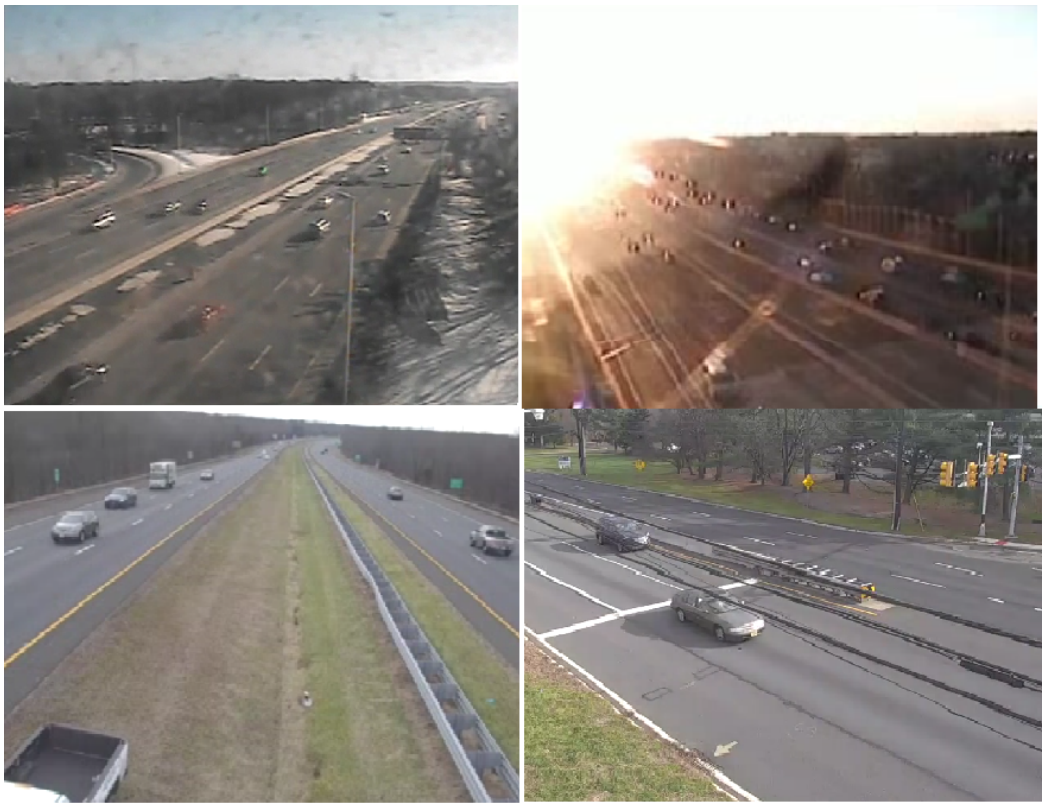


Figure 1.1 Some example frames from traffic videos showing low video quality.

cast shadows sometimes link two objects together and make the detected object's position inaccurate. Thus, the performance of video analysis is deteriorated by the cast shadows.

Region of Interest (RoI) detection is also lacking a mature solution. Road region is a widely used Region of Interest (RoI) in traffic video analysis. Figure. 1.3 shows some example video frames that the region of interest, which is the road region, only occupies a small portion of the frame. Detecting the road region ahead may help reduce the computation complexity of some video analysis applications, such as foreground detection and object detection because the algorithms do not need to process the areas outside the road.

Another topic people are also interested in is how to apply AI technologies to improve traffic safety. Driving in the correct direction is a primary rule to ensure the safety of the traffic. Therefore, detecting the anomaly in traffic becomes important in traffic video

Figure 1.2 Some example frames from different traffic videos showing strong cast shadows.

analysis. Figure. 1.4 shows a vehicle driving in the wrong direction, which is an extremely dangerous behavior that needs to be detected and stopped.

This dissertation, therefore, focuses on developing novel statistical modeling methods to solve these problems in traffic video analysis. First, a novel foreground detection method is presented, which includes a new Global Foreground Modeling (GFM) method, a Local Background Modeling (LBM) method, and the Bayes decision rule for minimum error. In the novel foreground detection method, a LBM method is first introduced, which is derived from the Gaussian mixture model. After that, an innovative GFM method is presented. The GFM method, which models the foreground pixels globally, is in contrast to the LBM method, which builds statistical models locally. Then the YIQ, the YCbCr, and some unconventional color spaces such as the Uncorrelated Color Space (UCS), the Independent Color Space (ICS), and the Discriminating Color Space (DCS) are investigated for feature extraction [67], [69]. Additionally, a novel feature vector, which integrates the color, wavelet, and temporal features, is introduced to increase the discriminatory power of the feature vector. Finally, the Bayes decision rule for minimum error is applied for final foreground and background classification.

Then, a new cast shadow detection method is presented for removing the cast shadows from the foreground detected by the foreground detection method. The novel moving cast shadow detection method contains four hierarchical steps. First, a set of new chromatic criteria is presented to detect the candidate shadow pixels in the HSV color space. We use the HSV color space for shadow detection due to its property of separating the chromaticity from intensity [29], [93], [21], [94], [38]. Our new chromatic criteria are more robust than the criteria used by other popular methods for shadow detection [94], [38]. Second, a new shadow region detection method is presented to cluster the candidate shadow pixels into shadow regions. Many shadow detection methods can not solve the shadow outlines problem: the outlines of the shadow regions are often classified to the foreground. As a result, after removing the shadow pixels from the foreground, the shadow regions are only partially removed, and the shadow outlines are often classified to the foreground. Our new shadow region detection method is able to solve this problem by applying the prior knowledge that both the foreground objects and their cast shadows should define continuous regions. Third, a statistical shadow modeling method, which uses a single Gaussian distribution to model the shadow class, is presented to classify shadow pixels. The shadow pixels detected by both the new chromatic criteria and the new shadow region detection method tend to be more reliable shadow pixels, therefore, these shadow pixels are used to estimate the Gaussian distribution for the shadow class. Finally, an aggregated shadow detection method is presented to integrate the detection results using the new chromatic criteria, the new shadow region detection method, and the new statistical shadow modeling method. A gray scale shadow map is obtained by calculating a weighted summation of the candidate shadow pixels. A shadow free foreground may be derived by thresholding the gray scale shadow map.

After that, a novel road detection method is introduced to solve the automated road recognition problem for the Region of Interest (RoI) detection in traffic video cognition. First, a temporal feature guided statistical modeling method is proposed for road modeling.

Figure 1.3 Some example frames showing that the road regions only occupy a small portion of the frame.

Specifically, a foreground detection method is applied to extract the temporal features from the video and then to estimate a background image. Furthermore, the temporal features guide the statistical modeling method to select sample data. Additionally, a model pruning strategy is applied to estimate the road model. Second, a new road region detection method is presented to detect the road regions in the video. The method applies discriminant functions to classify each pixel in the estimated background image into a road class or a non-road class, respectively. The presented method provides an intra-cognitive communication mode between the RoI selection and video analysis systems.

Furthermore, a novel anomalous driving behavior detection method is represented for traffic surveillance video analysis. First, a new Multiple Object Tracking (MOT) method is proposed to extract the velocities and trajectories of moving foreground objects in video. The new MOT method is a motion-based tracking method that integrates the temporal and spatial features. Second, a novel Gaussian Local Velocity (GLV) modeling method is presented to model the normal moving behavior in traffic videos. Note that the GLV model is learned and updated for every location in the video frame. Finally, a discriminant function is proposed to detect anomalous driving behaviors.

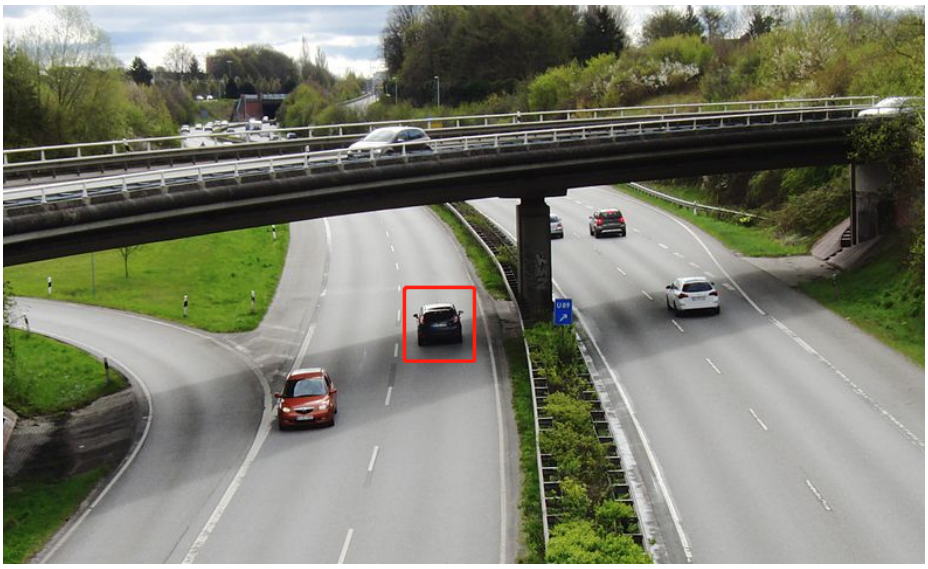


Figure 1.4 An example frame showing that a vehicle is driving in a anomalous direction.

This dissertation is organized as follows. Chapter 2 discusses some related work by other researchers on foreground detection, shadow removing, road recognition, and anomaly detection in video analysis. Chapter 3 explains the novel foreground detection method, which is able to achieve improved foreground detection performance and is capable of detecting stopped moving objects. Chapter 4 introduces the new cast shadow detection method, which is used to remove cast shadows from the foreground detection result and to enhance the video analysis performance. Chapter 5 represents the statistical modeling method for road recognition, which can automatically recognize the road regions in traffic videos as RoI. Chapter 6 represents an anomalous driving behavior detection method, which can detect the anomalies in traffic surveillance videos. Chapters 3, 4, 5, and 6 include detailed experimental results and comparisons between our presented methods with some popular methods using different video sets and show the advantages and capability of our methods. Chapter 7 summarizes our work and discusses the future work for research.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 Foreground Detection Methods

Many statistical modeling methods have been proposed for foreground detection [12], [17], [103], [14], [13]. Most of the statistical modeling methods build the models locally and are lacking accurate for foreground models [107], [136], [97], [9], [132], [13]. Wang [124] proposed a low-rank and sparse matrix method to detect foreground objects. Stauffer and Grimson [107] proposed a Gaussian Mixture Model (GMM) method and they do not apply any foreground modeling in their method. The GMM method builds one mixture Gaussian model for each location in a frame and uses a threshold to achieve the foreground and background classification. Based on the GMM method, Hayman and Eklundh further applied the Bayes classifier to separate the background and foreground by splitting the GMM into the background model and the foreground model, respectively [44]. The foreground model used by Hayman and Eklundh, which is the residual of the background, is determined locally. Zivkovic presented a background subtraction method that uses the GMM method to model the background and uses a uniform distribution to model the foreground for every location [136].

Note that these methods typically use the RGB colors as feature vectors, but the RGB color space sometimes cannot provide sufficient discriminating power for foreground detection. Further research indicated that some other color spaces, such as the YIQ and the YCbCr color spaces, are more powerful than the RGB color space in some visual tasks [130]. The YIQ and the YCbCr color spaces, which are broadly applied in the TV industry and in video and image compression [100], are defined by linear transformations of the RGB color space. Another defect of the RGB color space is that the RGB color values that are used as features are usually highly correlated [37], while most of the methods assume

that the feature vectors have independent features. The uncorrelated color space (UCS), the independent color space (ICS), and the discriminating color space [67], [69] are some innovated color spaces that can better satisfy this assumption. The UCS is derived from the RGB color space by using principal component analysis (PCA) [36]. The ICS uses the independent component analysis (ICA) [28], [54] to get three independent components. The DCS generates three new component images that are effective for classification by applying discriminant analysis [36].

As the feature vector that solely uses the color values often does not achieve sufficient discriminatory power, some foreground objects that have similar color values to the background may not be detected correctly. In order to increase the discriminatory power of the input feature vector, a number of region based methods are proposed [125], [86], [113], [91]. Wren *et al.* [125] presented a foreground object detection and tracking method based on blob-based features. Pandey and Lazebnik [86] used a multi-scale feature pyramid to present a video frame and trained a latent support vector machine (LSVM) for classification. Varadarajan *et al.* [113] proposed a region based foreground detection method, which uses small blocks as feature vectors to separate the foreground and the background. Qin *et al.* [91] used a background basis selection method that constructs the basis matrix of the background. The rationale of the region-based methods is to increase the discriminatory power of the input feature vector by increasing its size.

Another problem that is not adequately addressed by these foreground detection approaches is that the foreground objects cannot be detected when they stop moving, such as the vehicles that are stopped in traffic congestion. Li *et al.* [63] proposed a statistical modeling method in complex background environments for foreground object detection. Huang *et al.* [50] used a region-level motion-based background modeling method to detect foreground objects. They pointed out that the stopped foreground objects are always quickly absorbed by the background in their methods. As a result, these stopped moving targets are missed in video analysis.

2.2 Shadow Detection Methods

In video analysis, shadows are often detected as part of the foreground, as they share similar motion patterns to the foreground objects [89], [94], [75]. These cast shadows often adversely affect the video analysis performance in various applications, such as tracking and object detection. Hence shadow detection is a very important step to improve video analysis performance.

Many methods have been published for moving cast shadow detection [89], [94], [75]. As color often provides useful information for shadow detection, some methods apply color information to detect shadows [29], [21], [108], [4]. Many shadow detection methods assume that the shadow areas are darker in intensity but relatively invariant in chromaticity [29], [93], [21], [94], [38]. The color spaces that separate chromaticity from intensity are thus often used for shadow detection. Some example such color spaces are the HSV color space [29], the c1c2c3 color space [93], and the YUV color space [21]. Some popular methods apply a set of chromatic criteria by assuming that the cast shadows have a similar hue to the background, but a lower saturation and a lower value than the background [94], [38].

Statistical shadow modeling is applied for shadow detection as well [81], [49], [121]. The major assumption of these methods is that the light source is pure white and the attenuation of the illumination is linear. Generally speaking, these statistical shadow modeling methods are able to predict color changes of the shadow pixels better than the color based methods, but the shadow detection accuracy in outdoor scenes tends to deteriorate.

There are methods that use the shape, size, and orientation information for shadow detection [48], [34], [20]. These methods are designed to deal with some objects that have specific shapes. The advantage of these methods is that they do not need to estimate the background color of the shadow, but the disadvantage is that they have difficulty in dealing with multiple types of objects in complex scenes.

There are methods that utilize texture for shadow detection, such as classifying a region into the shadow region or the object region based on the texture correlation between the foreground and the background [126], [62], [95], [41], [115]. These methods extract the texture information in different sizes of the regions. The advantage of these methods is that they are more robust to illumination changes than the color based methods, but the disadvantage is that the computation efficiency of matching the texture features is low.

There are also methods that use machine learning techniques for shadow detection: a paired region based shadow detection algorithm is presented in [41], a kernel least-squares SVM method is proposed in [116] to separate shadow and non-shadow regions, and a shadow detection algorithm using a deep neural network is presented in [55]. These learning based shadow detection methods tend to have high computational complexity and are usually not for real time video analysis.

2.3 Road Region Detection Methods

Video surveillance cameras are widely deployed in modern society [85]. How to analyze surveillance videos becomes a very important topic. Traffic surveillance cameras are visible in our daily life. Automatic road recognition is an important task in analyzing traffic surveillance videos. Numerous approaches have been taken in order to segment the road region automatically.

The accumulation of the motion trajectories is widely used in road recognition. Melo *et al.* [78] modeled the motion trajectories of tracked vehicles by using low-degree polynomials and applied a K-means clustering technique on the coefficient space to obtain approximate lane centers. Lee *et al.* [61] generated a roadway mask image by accumulating moving parts in a difference map between two consecutive input frames and then identified a middle line of the roadway to separate two directions of the traffic. These approaches depend on the performance of the foreground detection and the tracking methods which can be affected by the quality of videos, changes in illumination, traffic density, etc.

Our proposed method, which utilizes the foreground detection result as guidance to build the statistical model, greatly reduces the inaccuracy caused by the foreground detection method.

Road markings and lane markings are the most commonly used features for road recognition [47], [70]. Wang *et al.* [123] introduced an algorithm called CHEVP to initialize a B-spline SNAKE algorithm and used the resulting B-spline curve to represent a curved road. Zhou *et al.* [134] estimated the lane model parameters (e.g., starting position, orientation, lane width, etc) and generated several lane model candidates and matched the best fitted lane model. Aly [3] proposed a real-time algorithm for detecting lane markings in urban streets by taking a top view of the road image, filtering with Gaussian kernels, line detection with Hough transformation, and a new RANSAC spline fitting approach. Kong *et al.* [57] used the OCR feature to estimate the vanishing point with a clustering method for road recognition. The vanishing point is estimated based on Gabor filters used to compute the dominant texture orientation at each pixel and a new edge detection technique to extract the road boundaries. Son *et al.* [104] proposed a real-time lane detection method to deal with illumination variations in lane departure warning system. After using the lane color properties to detect candidates for lane markers a clustering method was applied to find the main lane. Helala *et al.* [46] segmented the road into a number of superpixel regions and used the contours of these regions to generate several edges which are grouped into different clusters and the cluster with the highest confidence score was chosen as the road boundary. However, numerous roads have poor qualities that the markings are not clear or missing. It becomes difficult to locate the road when the road quality is poor or the resolution of the video is low.

Most recent studies try to use a combination of low-level and high-level features to deal with the road recognition problem, especially to overcome the effects of illumination changes and strong shadows. Wang *et al.* [119] introduced a close to real-time road recognition method based on illumination invariant image and quadratic estimation. After

extracting an illumination invariant image, a manual triangular road region was used as the color sample to analyze the illumination invariant image to obtain the probability maps. The combined probability map was resettled based on histogram analysis, and the road region was estimated for the first time. Then the effective road boundary was extracted after analyzing the gradient image by the estimated road region and the final more accurate road region was obtained. This method follows the assumption in [57] and [64] to use a manual triangular region that is approximated as the initial road estimation model. Tong *et al.* [112] used simple statistics to propose effective projection angle calculation methods in the logarithmic domain to extract the intrinsic images of roads in order to weaken the shadow effect and eliminate the impact of the direction of camera features. This method follows a similar approach based on [120] to use a prior triangle region to sample the color of the road region. Li *et al.* [65] proposed a road region extraction algorithm based on vanishing point location. The spatial structure of the road was estimated and color and edge features of the intrinsic image were extracted based on regression analysis.

Recently, many methods based on convolutional neural networks are proposed to solve the road detection problem [1], [19], [24], [60], [74], [88]. These methods that train the deep neural networks to segment the road regions need a huge amount of labeled training data, but still lack generalization ability. Their performance always drops significantly due to the impairments, such as different illumination and weather conditions, low image resolutions, and changes in camera viewing angles.

2.4 Anomalous Driving Detection Methods

Multiple Object Tracking (MOT) is one of the important steps in anomalous driving detection. Tracking multiple objects in video at the same time involves the detection of objects in each video frame and the association of the detected objects across multiple consecutive frames. There has been a significant amount of research conducted on Multiple Object Tracking (MOT) in recent years. Tang *et al.* [110] dealt with the MOT

problem as a Minimum Cost Subgraph Multicut Problem and applied the Kernighan-Lin algorithm to solve it. A fully differentiable graph-based framework and a Message Passing Network (MPN) to propagate the node features throughout the graph was proposed in [15]. Maksai *et al.* [76] used behavioral patterns to propose a non-Markovian approach in order to impose global consistency and further improved upon the state-of-the-art tracking algorithms. A deep prediction-decision network was developed by Ren *et al.* [92] using a collaborative deep reinforcement learning (C-DRL) method, which simultaneously detected and predicted objects under a unified network. With the improvements in object detection methods in recent years, tracking by detection has been the most studied approach in multi-target tracking. Kim *et al.* [56] modeled a multi-object state as a labeled random finite set, and used the Bayes recursion to reduce false negatives and false positives. The multi-object filtering density was propagated forward in time. A CNN-based framework was proposed by Chu *et al.* [27], which used single object tracking (SOT) to enrich detections in MOT. Jorquera *et al.* [53] used the Probability Density Hypothesis (PHD) filter and Determinantal Point Processes (DPP) to deal with data association uncertainty, noise, and false alarms and improved the detection accuracy.

Recently, anomaly detection in videos has been an active research area with applications in intelligent video surveillance and security related video analytic tasks. Since the occurrence of abnormal actions in real-world video analysis applications is infrequent, detecting anomalies in videos automatically reduces a significant amount of manual work. Some of the recent research efforts have tried to detect the anomalies in videos. A deep CNN was proposed by Nguyen *et al.* [83], which was a combination of a reconstruction network that determines the main structures that appear in video frames and an image translation model which associates motion templates to such structures. Liu *et al.* [72] proposed an anomaly detection method in videos with a video prediction framework, which predicted a future frame based on spatial and temporal constraints and then compared the predicted frame to its ground truth to detect abnormal incidents. An end-to-end network

was proposed by Tang *et al.* [111] that conducts future frame prediction. The network enlargers the reconstruction errors to help with the identification of abnormal events followed by reconstruction, which helps enhance the predicted future frames from normal events. Besides the methods that are aiming at detecting anomalies in general videos, many methods are proposed to detect anomalies in traffic videos. Doshi and Yasin proposed an unsupervised method to detect anomalies in traffic videos [32]. A three-stage pipeline for anomaly detection method was presented by [10].

CHAPTER 3

A NOVEL FOREGROUND DETECTION METHOD FOR VIDEO ANALYSIS BY INTEGRATING COLOR, WAVELET, AND TEMPORAL FEATURES

3.1 Introduction

Foreground detection, an important task in computer vision, is usually the first step in video analysis [12], [17], [35], [103], [14], [68], [99]. In order to detect the foreground objects, a common approach is to classify the foreground pixels and the background pixels into two classes.

Many statistical modeling methods have been proposed to address this problem, but the final solution remains elusive [107], [44], [136], [103], [59], [14], [98], [99]. First, the statistical modeling of the foreground is still lacking, such as assuming "uniform distribution for the foreground object appearance" [136] or using one Gaussian to model the background with the remaining Gaussians to model the foreground [44]. Second, the independence assumption among the component images in statistical modeling is not satisfied [98], [99]. Note that the three component images in the RGB color space are highly correlated. As a result, the statistical modeling methods that apply the RGB color space often yield sub-optimal, deteriorated foreground detection performance. Third, the discriminatory power of the input feature vector, which is defined by only the color information in a specific color space such as the RGB color space, is limited and inadequate for the complex task of foreground detection in video [98], [99].

In this chapter, we present a novel foreground detection method for video analysis to solve these problems. First, we present a new foreground detection method that consists of the Local Background Modeling (LBM) method, a novel Global Foreground Modeling (GFM) method, and the Bayes decision rule for minimum error. While the LBM models the background locally, the novel GFM method builds only one global model for the foreground. As a result, the GFM is a more accurate foreground modeling method when

compared to the published methods that assume uniform distribution or local foreground modeling [107], [44], [136], [9], [13]. The Bayes decision rule for minimum error finally classifies each pixel to the foreground class or the background class, respectively.

Second, in order to satisfy the independence assumption, we investigate the YIQ color space, the YCbCr color space, and some unconventional color spaces such as the Uncorrelated Color Space (UCS), the Independent Color Space (ICS), and the Discriminating Color Space (DCS) for feature extraction [67], [69].

Even though the RGB color space is the most commonly used color space in background subtraction methods [103], [14], [107], [44], [136], [59], its three color component images are highly correlated and the independence assumption is not satisfied [37].

The YIQ and the YCbCr, which are broadly used in the TV industry and in video and image compression [100], are able to separate the chromatic and the achromatic values. As a result, these color spaces are better choices than the RGB color space.

The UCS, DCS, and ICS, which are derived using the principal component analysis or PCA, discriminant analysis, and the independent component analysis or ICA, all have uncorrelated color component images. Furthermore, the ICS satisfies the independence assumption due to its independent color component images.

Third, to further enhance the discriminatory power of the input feature vector, we augment the three-dimensional feature space to define a new 12-dimensional feature vector space by integrating the horizontal and vertical Haar wavelet features [117], [66] and the temporal information into the color features. Note that the new color spaces help reduce the correlation in the RGB color space, and the ICS is able to derive three color components that are statistically independent. In addition, the 12-dimensional feature vector is able to increase the discriminatory power [68], which helps our foreground detection method achieve better performance than other popular statistical modeling methods.

We implement experiments using the videos from the New Jersey Department of Transportation (NJDOT) and the public data set CD.net-2014 [122], and compare the performance of our proposed foreground detection method with some other popular statistical modeling methods in different feature spaces [107], [44], [136], [137], [106]. We also show that our proposed method can detect the temporarily stopped moving foreground objects in video, such as vehicles stopping in front of traffic lights.

3.2 A Novel Foreground Detection Method for Video Analysis

The foreground detection problem is in essence a two-class pattern classification problem, since if we can correctly classify each pixel into a foreground class or a background class, we are able to detect the foreground objects. Even though many statistical modeling methods have been proposed to address the foreground detection problem in video [107], [44], [136], [103], [59], [14], some challenging problems are still waiting for satisfactory solutions: (i) more accurate foreground modeling, (ii) the independence assumption among the color component images, and (iii) the inadequate discriminatory power of the input feature vector.

To address these challenging problems, we propose a novel foreground detection method that includes a Local Background Modeling method (LBM), a novel Global Foreground Modeling (GFM) method, color spaces that better satisfy the independence assumption, and a new 12-dimensional feature vector space that enhances the discriminatory power of the GFM method.

Specifically, first, the GFM method estimates a global probability density function for the foreground and applies the Bayes decision rule for model selection, while the LBM method selects the most significant Gaussian distribution from the Gaussian mixture model locally as the background. Second, to mitigate the high correlation effects of the RGB color space on the independence assumption among the color component images, we investigate the YIQ, the YCbCr, and some unconventional color spaces such as the Uncorrelated Color

Space (UCS), the Independent Color Space (ICS), and the Discriminating Color Space (DCS) for feature extraction [67], [69]. Note that the YIQ and the YCbCr separate the chromatic and the achromatic values, the UCS and DCS further decorrelate the component images, and the ICS satisfies the independence assumption by deriving the independent component images. Third, to further enhance the discriminatory power of the feature vector, we augment the three-dimensional feature space to define a new 12-dimensional feature vector space by integrating the horizontal and vertical Haar wavelet features [117], [66] and the temporal information into the color features.

3.2.1 A Local Background Modeling (LBM) Method Using a Single Gaussian Density

For background modeling, one single Gaussian density is learned in the novel feature vector space $\mathbb{R}^{12}$ (see 3.2.5) to model the background locally for each pixel. Specifically, the local background modeling involves a two-step process. First, the probability density function of the pixel at (i, j) is estimated using the traditional local Gaussian Mixture Model (GMM) [136], [44], [107]. The constant weight updating scheme [80] is then applied to learn the parameters of the GMM. Second, according to the weights of the Gaussian densities, we choose the most significant single Gaussian density to model the background locally for each pixel.

The feature vector, $\mathbf{x}_{i,j} \in \mathbb{R}^{12}$, is defined by the three color values of the pixel, the horizontal and vertical Haar wavelet features, and the temporal difference features. For notational simplicity and without loss of generality, we will drop the subscripts i, j in the following equations. The probability density function of the pixel at (i, j) may be estimated as follows [107]:

$$p(\mathbf{x}) = \sum_{l=1}^{L} \alpha_l N(\mathbf{M}_l, \Sigma_l) \tag{3.1}$$

$$N(\mathbf{M}_l, \Sigma_l) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{M}_l)^t \Sigma_l^{-1} (\mathbf{x} - \mathbf{M}_l)\right\}}{(2\pi)^{d/2} \mid \Sigma_l \mid^{1/2}} \tag{3.2}$$

$$\sum_{l=1}^{L} \alpha_l = 1 \tag{3.3}$$

where L indicates the number of Gaussian densities in the GMM model, α_l is the weight for each Gaussian density, d is the dimensionality of the feature vector $\mathbf{x}$, and $\mathbf{M}_l$ is the mean vector and Σ_l is the covariance matrix for the l-th Gaussian density.

For simplicity, we assume that the covariance matrix Σ_l is diagonal. Note that the traditional pixel-based background subtraction algorithms use the same assumption as well [136], [44], [107]. Generally speaking, the GMM is a comprehensive model for describing complex scenes with various activities. Thus both the background and the various activities are described by the different Gaussian densities.

Note that the background, which is usually static without many changes, may be modeled by one Gaussian density with a large weight. We thus choose the most significant Gaussian density, which is the first density in the GMM, to model the background.

$$p(\mathbf{x}|\omega_b) = N\left(\mathbf{M}_1', \Sigma_1'\right) \tag{3.4}$$

where ω_b represents the background class.

3.2.2 A Novel Global Foreground Modeling (GFM) Method

Our novel Global Foreground Modeling (GFM) method, which differs from the LBM method that builds models locally at each location, constructs a global foreground model for all the foreground pixels in the whole video frame. Note that a moving foreground object can exist at any location in the region of interest (ROI). For example, a vehicle can move along a road. We first use all the foreground information to learn K Gaussian density functions in order to model the foreground objects. Then for each location, we apply the Bayes decision rule for minimum error to choose one learned Gaussian density function as the foreground model.

Specifically, we define K Gaussian density functions as the conditional density functions for all the foreground pixels in the GFM method, respectively: $p(\mathbf{x}|\omega_1)$, $p(\mathbf{x}|\omega_2)$, $\cdots$, $p(\mathbf{x}|\omega_K)$, which are defined as follows:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k p(\mathbf{x}|\omega_k) \tag{3.5}$$

$$p(\mathbf{x}|\omega_k) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{M}_k)^t \Sigma_k^{-1}(\mathbf{x}-\mathbf{M}_k)\right\}}{(2\pi)^{d/2} \mid \Sigma_k \mid^{1/2}} \tag{3.6}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \tag{3.7}$$

where $\mathbf{x}$ presents the feature vector with a dimensionality of d, $\mathbf{M}_k$, Σ_k, and α_k are the mean vector, the covariance matrix, and the weight for the k-th Gaussian density, respectively. For each density $p(\mathbf{x}|\omega_k)$, we define a counter n_k for counting the number of input feature vectors.

Next, we initialize the foreground model. In particular, we apply K Gaussian distributions for the foreground model, and each distribution is associated with a counter and a weight. For each pixel, we define a new feature vector $\mathbf{x}$ that contains the color components, the horizontal and vertical Haar wavelet features [117], [66], and the temporal difference features (see Subsection. 3.2.5 for details). The elements of the mean vector and the covariance matrix of each Gaussian distribution are first initialized to zero, and so are the corresponding counter and weight.

Before all the K Gaussian distributions are defined, if $\mathbf{x}$ is within the threshold of the corresponding background density, we classify it as a background pixel and do not modify the foreground model. Otherwise, if it is within the threshold of any existing foreground Gaussian distribution, we use that feature vector to update the parameters in the corresponding Gaussian distribution. The updating strategy is as follows [80]:

$$\mathbf{M}_k' = \frac{n_k \mathbf{M}_k + \mathbf{x}}{n_k + 1} \tag{3.8}$$

$$\Sigma_k' = \frac{n_k \Sigma_k + (\mathbf{x} - \mathbf{M}_k)(\mathbf{x} - \mathbf{M}_k)^t}{n_k + 1} \tag{3.9}$$

$$n_k' = n_k + 1 \tag{3.10}$$

$$\alpha_k' = \frac{n_k'}{\sum_{k=1}^{K} n_k'} \tag{3.11}$$

where $\mathbf{M}_k$ and Σ_k are the mean vector and covariance matrix for the k-th Gaussian distribution with the corresponding counter n_k and the weight α_k. If the feature vector is not within the threshold of any existing foreground Gaussian distribution, we create a new Gaussian density function using $\mathbf{x}$ as the mean vector and a predefined value σ_0 as the diagonal values in the covariance matrix. We then set the corresponding counter to one and update the weights of all the K Gaussian distributions using Equation (3.11).

After all the K Gaussian distributions are defined, we apply the Bayes decision rule for minimum error [5] to assign the input feature vector to one Gaussian distribution.

$$p(\mathbf{x}|\omega_f)P(\omega_f) = \max_{i=1}^{K}\{p(\mathbf{x}|\omega_i)P(\omega_i)\} \tag{3.12}$$

The Gaussian distribution $p(\mathbf{x}|\omega_f)$ will be used as the foreground conditional probability density function and $\mathbf{x}$ will be used to update the parameters for $p(\mathbf{x}|\omega_f)$. Note that ω_f represents the foreground class.

3.2.3 Foreground and Background Classification

Finally, we can apply the Bayes decision rule for minimum error to classify the pixels into a background class and a foreground class.

For a specific location (i, j) at time t, the conditional probability density function (CPDF) for the foreground and the background, $p(\mathbf{x}|\omega_f)$ and $p(\mathbf{x}|\omega_b)$, are defined in the previous sections, respectively. Note that we assume that the background is a static scene without significant motion, and one Gaussian distribution is sufficient to describe the

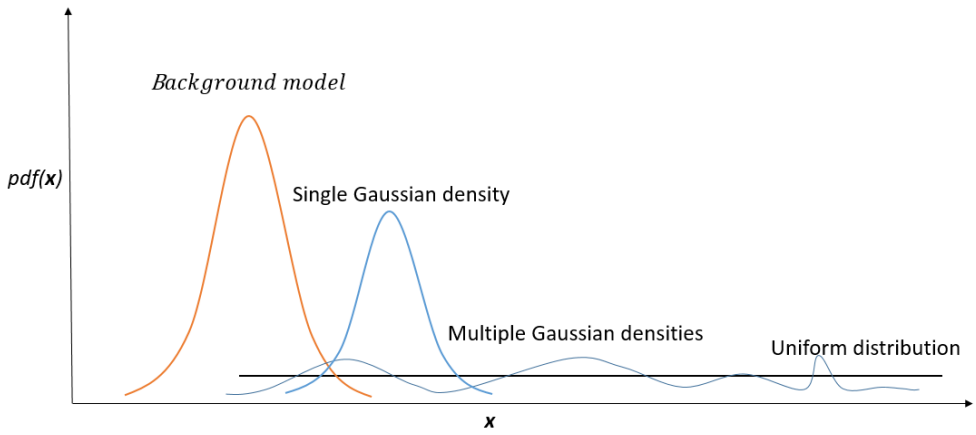


Figure 3.1 A comparison between a single Gaussian density with multiple Gaussian densities and a uniform distribution in foreground detection. The single Gaussian density has a smaller overlapping with the background model, which means a lower error rate.

background information. As discussed in Subseection 3.2.1, the first Gaussian density is chosen as the background model, the prior probability for the background, $P(\omega_b)$, is estimated using the weight for this Gaussian density: $P(\omega_b) = \alpha_1$. The prior probability for the foreground, $P(\omega_f)$, is then estimated as follows: $P(\omega_f) = 1 - \alpha_1$.

Using the statistical functions estimated above, we can classify each pixel in a video frame into the background or the foreground class. Given a pixel, we first convert it into a specific feature space, the RGB, the YIQ, the YCbCr, the UCS, the ICS, the DCS, or the 12-dimensional feature spaces and obtain the input feature vector $\mathbf{x}$. We then apply the Bayes decision rule for minimum error by means of the following discriminant function:

$$c(\mathbf{x}) = p(\mathbf{x}|\omega_f)P(\omega_f) - p(\mathbf{x}|\omega_b)P(\omega_b) \tag{3.13}$$

The feature vector $\mathbf{x}$ is classified to the foreground class if $c(\mathbf{x}) > 0$, and to the background class otherwise. Therefore, we can detect the foreground objects in the original frame according to the pixels belonging to the foreground class.

Our novel foreground detection method has the following advantages. First, our GFM method estimates a global statistical model which includes all the foreground information,

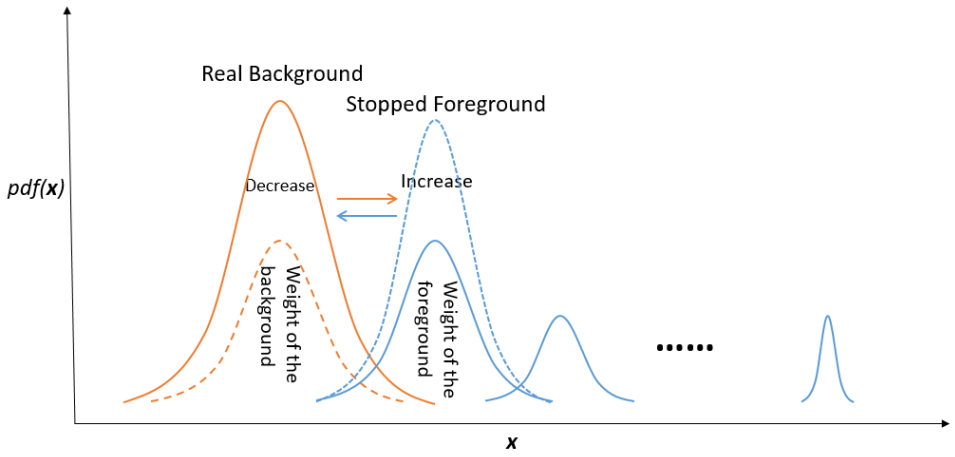


Figure 3.2 The weight of the background Gaussian density will decrease when an object stops moving. The background model will be replaced with a new Gaussian distribution which represents the stopped moving object.

and applies the Bayes decision rule for minimum error to select one single Gaussian density for foreground classification. In contrast to Hayman and Eklundh's method [44] and Zivkovic's method [136], which use the residual of the GMM background model or a uniform distribution to model the foreground, our global foreground model is more accurate for foreground modeling. Figure 3.1 compares a single Gaussian distribution with multiple Gaussian distributions and a uniform distribution in foreground detection. We can see that the single Gaussian distribution has a smaller overlapping with the background model, which means a smaller error. Therefore, our novel foreground model is able to achieve better foreground detection performance.

Second, our global foreground detection method can detect foreground objects that stop moving temporarily. Many popular methods fail to detect stopped moving objects [107], [44], [63], [136], [50], [103], [14]. The main reason is that these methods do not have an accurate foreground model. When an object stops moving, the background model will be replaced with a new Gaussian distribution which represents the stopped moving object. Figure 3.2 shows the change of the background model when a foreground object stops moving. For Zivkovic's method, as the foreground is modeled by uniform

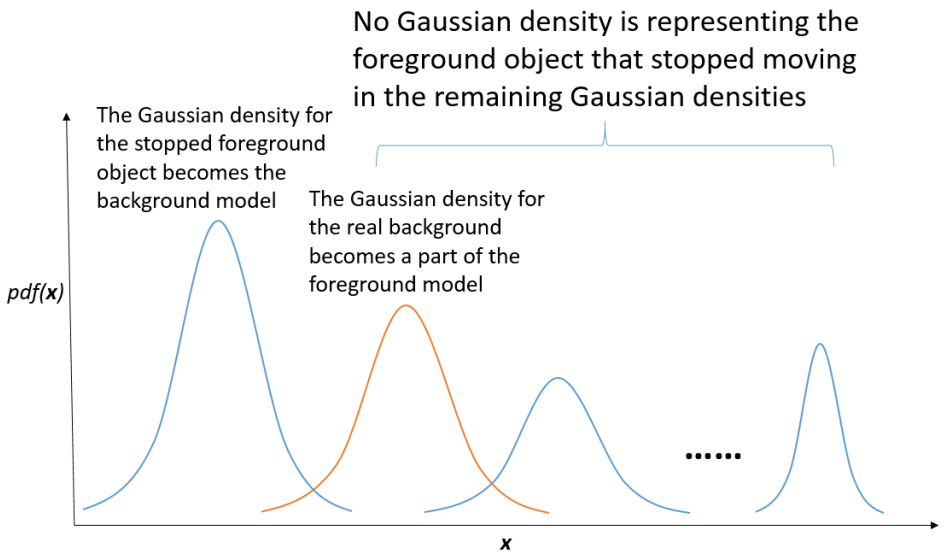


Figure 3.3 In Hayman and Eklundh's method, after an object stops moving, the foreground model will no longer remain any information of that foreground object.

distribution, there is no foreground information stored in that model. For Hayman and Eklundh's method, when a moving object stops, the background Gaussian distribution for each location corresponding to that object becomes part of the foreground model and the Gaussian distribution representing that object becomes the background model. We can see from Figure 3.3, the foreground model will no longer retain any information of that foreground object. So the stopped foreground object cannot be detected with these models. Note also that these methods use the exponential updating scheme to update the background model, which will cause the background model to change very fast when an object stops.

Our GFM method, however, maintains a foreground model that is not relevant to the background model. Even the background model may be influenced by the stopped moving object, our foreground model will keep the correct foreground object information. When applying the Bayesian decision rule, our method has a better chance to classify a pixel into its correct class. In addition, we apply the constant weight updating scheme to update the background model, which can also reduce the influence of the stopped moving object.

3.2.4 Feature Vectors in Various Color Spaces

Color is broadly applied for feature extraction in computer vision and pattern recognition [71], [90], [68]. The RGB color space, which contains the red, green and blue color component images, is a commonly used color space in foreground detection methods [107], [44], [136], [99], [98]. One common assumption in the statistical modeling methods is that the color components in the feature vector are independent [107], [44], [136]. This assumption simplifies statistical modeling because the independent variables lead to a diagonal covariance matrix. However, this assumption is not satisfied when the RGB color space is used, as the red, green, and blue color component images are highly correlated hence not independent in the RGB color space [37]. As a result, the statistical modeling in the RGB color space by assuming the independence of the color components will not produce the optimal foreground detection performance.

In order to better satisfy the independence assumption, we apply additional color spaces, such as the YIQ color space, the YCbCr color space, the uncorrelated color space (UCS), the independent color space (ICS), and the discriminating color space (DCS) [67], [69], [101] for improving the foreground detection performance.

The RGB color space is the most widely used color space in foreground detection methods [107], [44], [136], [14], [103] which has highly correlated components [37]. The YIQ and YCbCr color spaces separate the chromatic and the achromatic values. The UCS is derived by de-correlating the three component images in the RGB color space using principal component analysis (PCA) [36]. The ICS, which applies independent component analysis (ICA) [28], [54] can further enhance the discriminating power for our foreground detection task. Furthermore, the ICS derives three color components that are statistically independent, which satisfies the independence assumption. The DCS generates three new component images that are effective for classification by applying discriminant analysis [36]. The rationale of applying these color spaces is to mitigate the correlation effects of the RGB color space on the independence assumption among the color component images.

The YIQ color space and the YCbCr color space are defined as follows [39]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5957 & -0.2745 & -0.3213 \\ 0.2115 & -0.5226 & 0.3111 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.14}$$

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.2568 & 0.5041 & 0.0979 \\ -0.1481 & -0.2908 & 0.4390 \\ 0.4391 & -0.3677 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.15}$$

For simplicity, we apply the learned transformation matrices in [67] to define the UCS, ICS, and DCS for statistical modeling. In particular, the transformation matrices for the UCS, the ICS, and the DCS (W_U, W_I, and W_D) are as follows [67]:

$$W_U = \begin{bmatrix} 0.8836 & 0.3660 & 0.2922 \\ -0.4574 & 0.5411 & 0.7057 \\ -0.1002 & 0.7572 & -0.6455 \end{bmatrix} \tag{3.16}$$

$$W_I = \begin{bmatrix} -2.3286 & 1.1997 & 0.9774 \\ 1.2906 & -0.7658 & -2.1713 \\ 0.4125 & -1.2598 & 0.9212 \end{bmatrix} \tag{3.17}$$

$$W_D = \begin{bmatrix} -0.4258 & 0.7918 & -0.4378 \\ 0.0440 & 0.5548 & -0.8308 \\ 0.1985 & -0.9019 & 0.3835 \end{bmatrix} \tag{3.18}$$

Figure 3.4 shows the color component images in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The first row shows the R, G, and B component images in the RGB color space, the second row shows the Y, I, and Q component images in the YIQ color space, the third row shows the

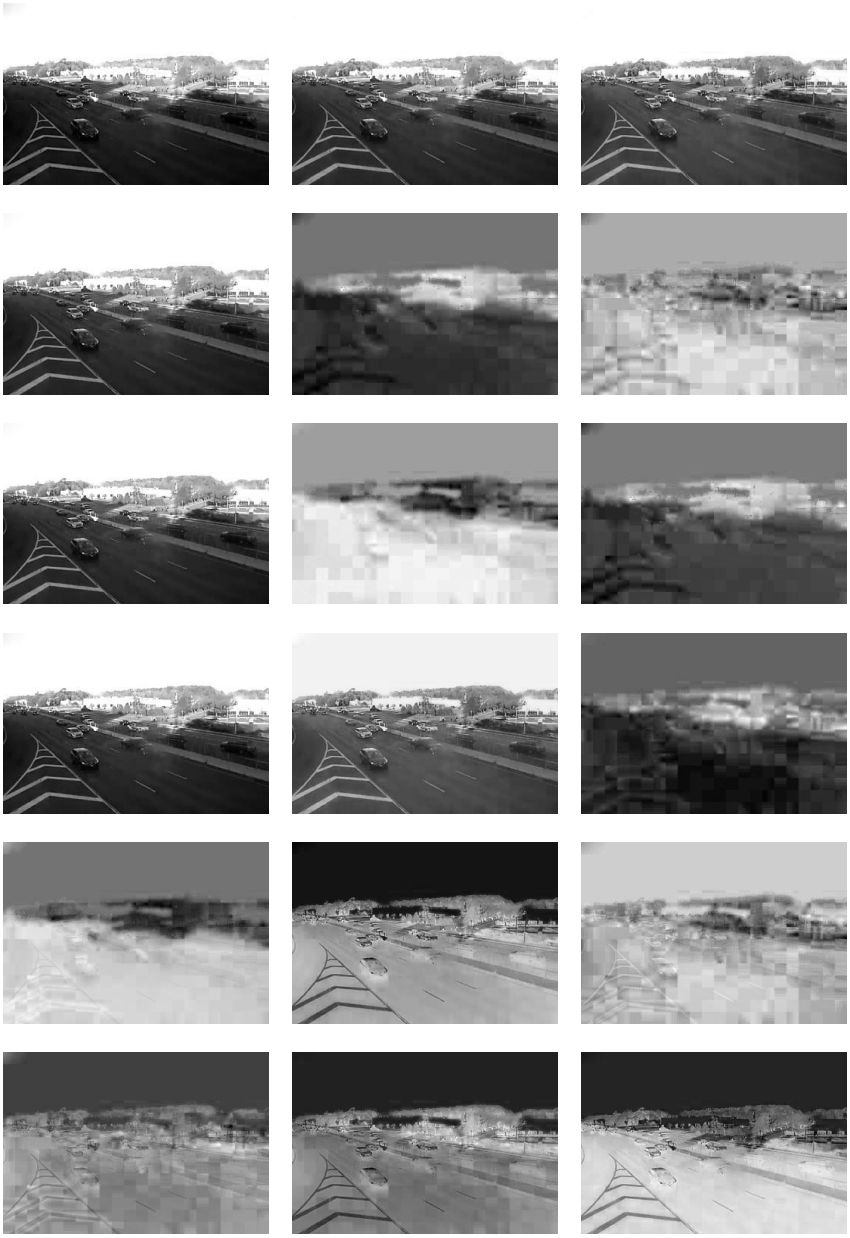


Figure 3.4 Color component images in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The first row shows the R, G, and B component images in the RGB color space, the second row displays the Y, I, and Q component images in the YIQ color space, the third row displays the Y, Cb, and Cr component images in the YCbCr color space, the fourth row shows the three uncorrelated component images in the UCS, the fifth row shows the three independent component images in the ICS, and the sixth row shows the three discriminating component images in the DCS.

Y, Cb, and Cr component images in the YCbCr color space, the fourth row shows the three uncorrelated component images in the UCS, the fifth row shows the three independent component images in the ICS, and the sixth row shows the three discriminating component images in the DCS.

Applying the color information solely for foreground detection in videos sometimes is inadequate, as evidenced by the fact that the color values of the foreground and the background sometimes are quite similar. To further enhance the discriminatory power of the feature vectors, we propose to augment the color feature vector with additional discriminatory information, such as wavelet features and temporal information.

3.2.5 Enhancing Discriminatory Power of the Feature Vector by Integrating Color, Wavelet, and Temporal Features

Feature representation plays a very important role in pattern classification [68], [71], [90], [23], [102], [6]. Our recent research shows that the discriminatory power of the feature vector is enhanced by increasing the dimensionality of the feature vector [22]. The popular background subtraction algorithms usually apply the red, green, and blue values of a pixel to define the input vector [14], [103]. As a result, the size of the input vector is limited, which restricts the discriminatory power of the vector.

To enhance the discriminatory power of the feature vector, we augment the three-dimensional feature space to define a new 12-dimensional feature vector space by integrating the color features, the horizontal and vertical Haar wavelet features [117], [66], as well as the temporal information in the video.

Specifically, first, the new feature vector incorporates the color values of a pixel. The color values are useful for some simple segmentation tasks as demonstrated in the paper [42]. Second, the new feature vector integrates the horizontal and vertical Haar wavelet features. Haar wavelet features have been broadly applied in computer vision and pattern recognition [66]. Third, the new feature vector combines the temporal difference features. As temporal information plays a crucial role in motion analysis, the temporal difference

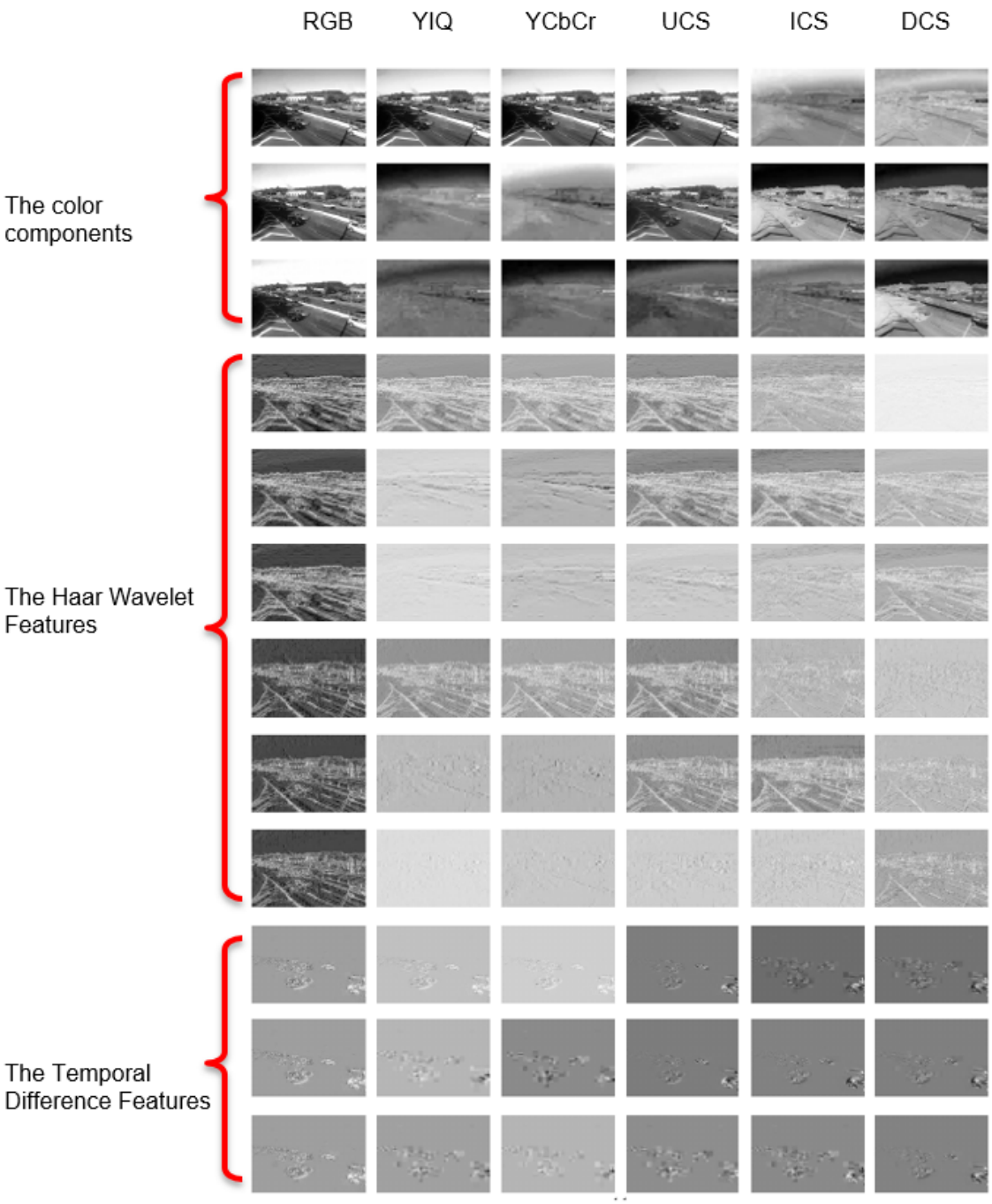


Figure 3.5 The feature vectors in different color spaces. Each column shows the 12-dimensional features in one color space, namely, the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. Each column contains the three color components, the horizontal Haar wavelet features, the vertical Haar wavelet features, and the temporal difference features in three corresponding colors.

Table 3.1 The Run Time of the Foreground Detection Method using the Three Types of the NJDOT Traffic Videos: 352×240 Video, 704×480 Video, and 752×480 Video

video quality	resolution	run time
high quality	752×480	82ms/frame
enhanced quality	704×480	68ms/frame
existing quality	352×240	15ms/frame

features are added to our new feature vector. Our idea is to compute the temporal difference between the current frame and the next frame with the goal of distinguishing the moving objects and the stable background.

In Figure 3.5, each column shows the 12-dimensional features in one color space, namely, the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. Each column contains the three color components, the horizontal Haar wavelet features, the vertical Haar wavelet features, and the temporal difference features in three corresponding color components.

3.3 Experiments

We analyze in this section the foreground detection performance of our proposed method using video sequences from the New Jersey Department of Transportation (NJDOT), and the public data set CD.net-2014 [122]. The NJDOT video sequences are real video data recorded by traffic surveillance cameras. In particular, the NJDOT data set contains various videos with different spatial resolutions and frame rates, such as 352×240 with 15 frames per second (FPS), 704×480 with 15 fps, 752×480 with 30 fps. We use several dozens of the NJDOT videos to evaluate the foreground detection performance of our method and the experimental results show that our proposed method is able to achieve real-time processing of the videos. The CD.net-2014 data set is a well known public data set for foreground detection. The dataset contains several simulated video sequences by using the separate frames, and provides pixel wised foreground mask of each frame for evaluation.

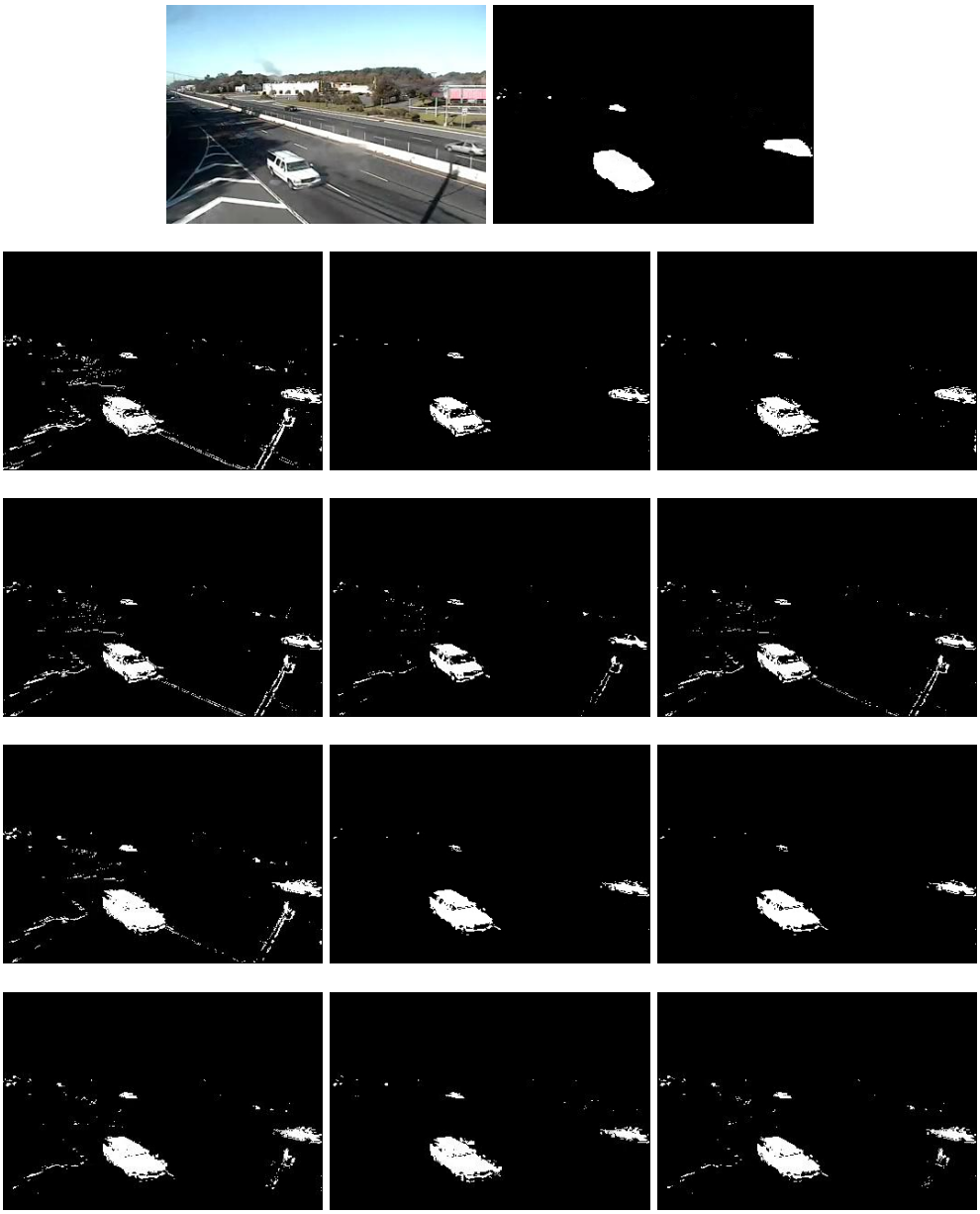


Figure 3.6 Comparative foreground detection performance of the proposed foreground detection method in three-dimensional feature paces and 12-dimensional feature spaces. The First row shows one video frame from an NJDOT traffic video with spatial resolution of 352×240 and the ground truth of the foreground mask. The second and the third rows show the foreground detection results using the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The fourth and the fifth rows show the foreground detection results using the 12-dimensional feature vectors constructed from the corresponding color spaces, respectively.

We choose to use the baseline videos in the CD.net-2014 to evaluate our proposed method and compare the performance with some other recently published unsupervised methods.

We used a DELL XPS 8900 desktop PC with an Intel Core i7-6700 Processor to implement our global foreground detection method. The parameters we used are as follows: the number of Gaussian density functions for the GMM is 3, and the number of Gaussian density functions for the GFM is 5. Table. 3.1 shows the running time of our method on different videos. These experimental results indicate that our proposed method is able to process the existing quality videos and the enhanced quality videos in real time on the DELL XPS 8900 PC with a 3.4 GHz processor. As a matter of fact, the existing quality videos are currently in use by NJDOT for traffic monitoring. The other two types are collected using temporary cameras at some experiment zones.

We first evaluate the foreground detection performance of our foreground modeling method in the proposed 12-dimensional feature spaces corresponding to the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. Note that for a specific color space, the 12-dimensional feature vector as shown in Figure 3.5 is defined by the three values of the three color component images and the nine values of the horizontal, the vertical Haar wavelet features as well as the temporal difference features of the corresponding color component images, respectively. The foreground detection performance using different feature vectors and our proposed foreground detection method is shown in Figure 3.6.

In particular, the first row in Figure 3.6 displays one traffic video frame from the NJDOT and the ground truth of the foreground mask. The second and third rows show the foreground detection results using the three-dimensional feature vectors in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. The fourth and the fifth rows show the foreground detection results using the 12-dimensional feature vectors constructed from the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively.

Table 3.2 The Foreground Detection Performance of Our Proposed Foreground Detection Method using Different Feature Vectors

3-D	**RGB**	**YIQ**	**YCbCr**	**UCS**	**ICS**	**DCS**
Precision	58%	97%	91%	65%	81%	70%
Recall	67%	61%	70%	62%	52%	58%
F-measure	62%	75%	79%	63%	63%	63%
12-D	**RGB**	**YIQ**	**YCbCr**	**UCS**	**ICS**	**DCS**
Precision	71%	95%	95%	80%	89%	84%
Recall	87%	76%	74%	84%	82%	82%
F-measure	78%	84%	83%	82%	**85%**	83%

Figure 3.6 reveals that the foreground detection results in the 12-dimensional feature spaces are better than those in the three-dimensional feature spaces. To quantitatively measure the foreground detection performance of the experimental results in Figure 3.6, we apply the broadly used metrics precision, recall, and the F-measure, which are defined as follows [40]:

$$Precision = \frac{TP}{TP + FP} \tag{3.19}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.20}$$

$$F - measure = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{3.21}$$

where TP, FP, and FN represent the number of true positive, false positive, and false negative foreground pixels, respectively.

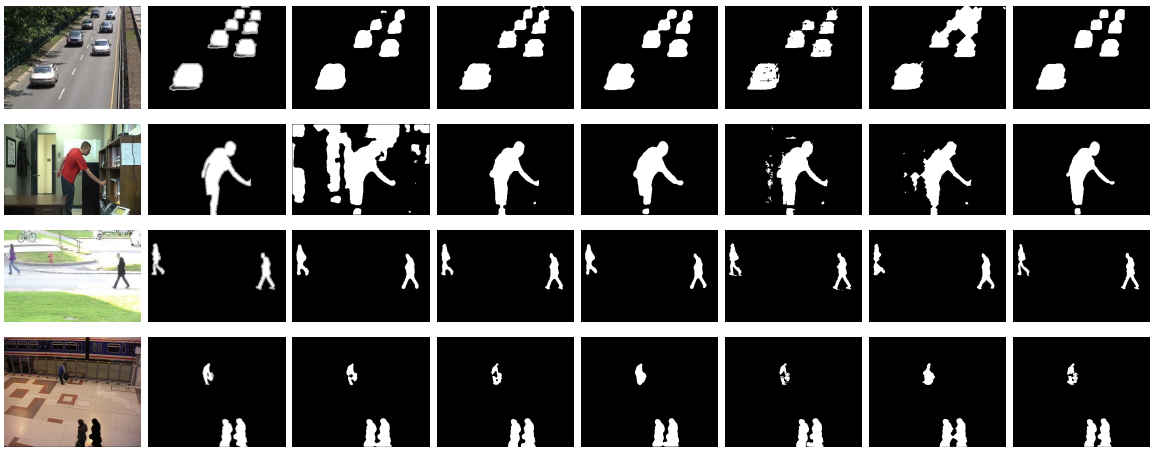


Figure 3.7 Comparative performance of the proposed foreground detection method and other popular methods using the CD.net-2014 videos. Each row represents a different video. From top to bottom are highway, office, pedestrians, and PETS2006, respectively. The First column shows some video frames from the data set. The second column shows the ground truth. The third to the seventh columns display the detection results using the BMOG method [77], the CL-VID method [73], the PAWCS method [106], the SBBS method [114], and the SWCD method [51], respectively. The last column displays the detection results of our proposed method in the ICS.

Table 3.2 shows the precision, recall, and the F-measure scores of the experimental results in Figure 3.6 corresponding to the three-dimensional feature spaces and the 12-dimensional feature spaces, respectively.

Specifically, the F-measure scores indicate that (i) the 12-dimensional feature vectors outperform the three-dimensional feature vectors for foreground detection in video, (ii) the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS outperform the traditional RGB color space for foreground detection in video, and (iii) the ICS, which satisfies the independence assumption among the color component images, achieves the best foreground detection in video.

We then compare our proposed method with some popular foreground detection methods using the CD.net-2014 videos. Figure 3.7 presents the comparative foreground detection performance of the BMOG method [77], the CL-VID method [73], the PAWCS method [106], the SBBS method [114], the SWCD method [51], and the proposed

foreground detection method using the CD.net-2014 videos. We also quantitatively compared our proposed method with these methods. From Table. 3.3, we can see that our proposed method reaches the state-of-art foreground detection accuracy.

Table 3.3 The Foreground Detection Performance of Different Methods using the CD.NET-2014 Videos

	Method	**BMOG**	**CL-VID**	**PAWCS**	**SBBS**	**SWCD**	**Proposed**
highway	Precision	94%	92%	94%	93%	87%	**94%**
	Recall	95%	95%	74%	90%	95%	**97%**
	F-measure	95%	95%	94%	92%	91%	**95%**
office	Precision	74%	96%	97%	96%	91%	**99%**
	Recall	55%	95%	91%	**97%**	**97%**	93%
	F-measure	63%	95%	94%	**97%**	94%	96%
pedestrians	Precision	87%	91%	93%	89%	90%	**96%**
	Recall	98%	**99%**	**99%**	83%	96%	92%
	F-measure	92%	**95%**	**95%**	94%	93%	94%
PETS2006	Precision	74%	85%	**92%**	81%	86%	89%
	Recall	94%	94%	94%	91%	**96%**	89%
	F-measure	83%	89%	**93%**	86%	91%	89%
Average	Precision	82%	91%	94%	90%	89%	**95%**
	Recall	86%	**97%**	94%	94%	96%	93%
	F-measure	83%	94%	94%	92%	92%	**94%**

In addition, we compare our proposed method with some popular foreground detection methods, namely, the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the Pixel-Based Adaptive Word Consensus Segmenter (PAWCS) method [106] for foreground detection using the NJDOT traffic videos. Figure 3.8 shows the comparative foreground detection performance using these

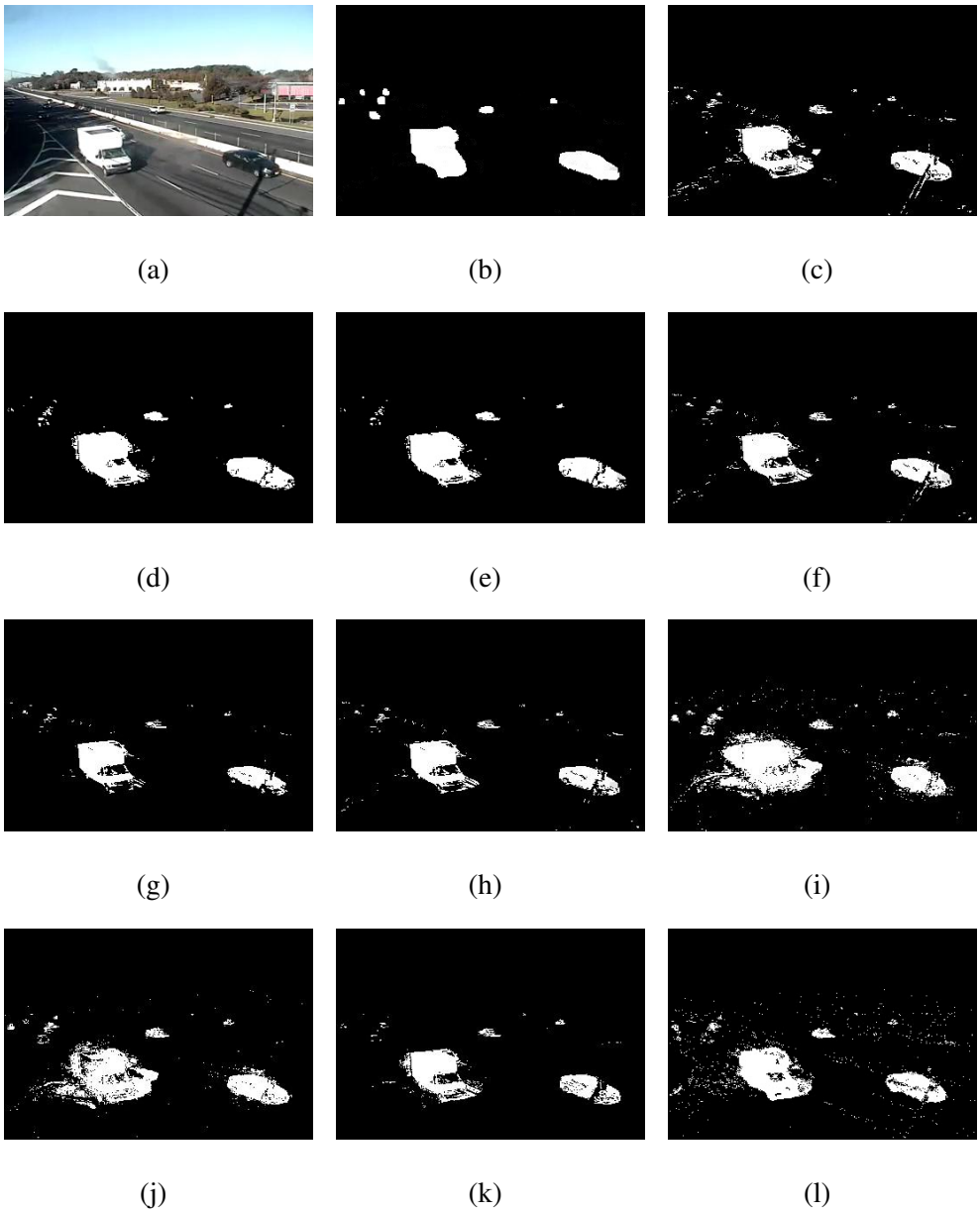


Figure 3.8 Comparative foreground detection performance of the proposed foreground detection method and some popular video analysis methods. (a) One video frame from an NJDOT traffic video with a spatial resolution of 352×240. (b) The ground truth of the foreground mask. (c)-(h) The foreground masks that are extracted using the proposed method in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. (i)-(l) The foreground masks which are extracted using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106], respectively.

Table 3.4 The Foreground Detection Performance of Different Methods

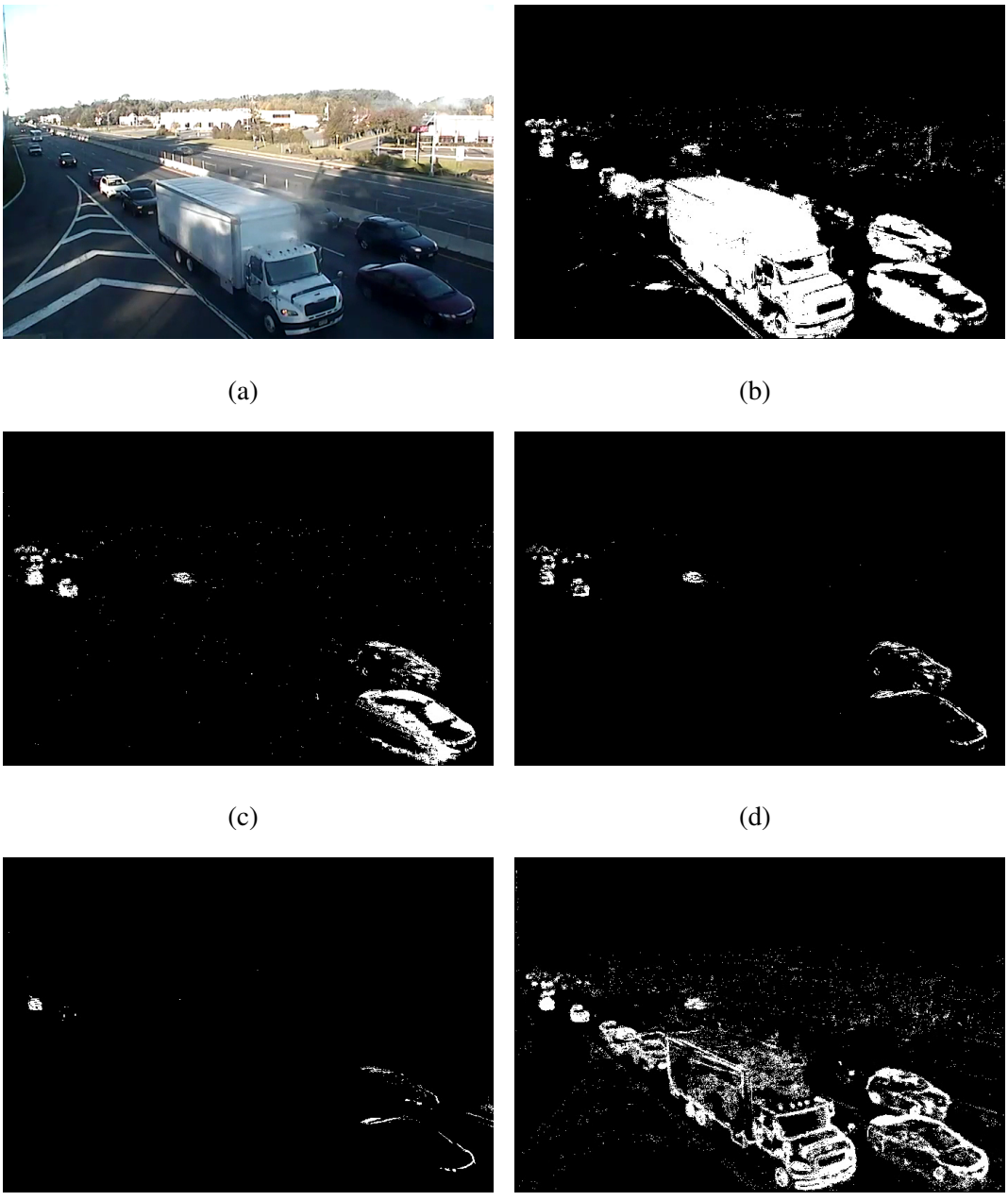
Method	Precision	Recall	F-measure
GMM [107]	54%	86%	66%
Zivkovic's [136], [137]	59%	82%	69%
Hayman and Eklundh's [44]	83%	74%	78%
PAWCS [106]	68%	83%	74%
Proposed Method in ICS	89%	82%	**85%**

methods. In particular, Figure 3.8 (a) shows a video frame from an NJDOT traffic video with spatial resolution of 352×240. Figure 3.8 (b) shows the ground truth of the foreground mask. Figure 3.8 (c)-(h) show the foreground detection results using our proposed method in the RGB color space, the YIQ color space, the YCbCr color space, the UCS, the ICS, and the DCS, respectively. Figure 3.8 (i)-(l) show the foreground detection results using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106].

The comparative foreground detection performance in Figure 3.8 reveals that our proposed method may achieve better foreground detection results than the popular foreground detection methods. The precision, recall, and the F-measure scores of the experimental results in Figure 3.8 are shown in Table. 3.4.

Specifically, Table. 3.4 compares the foreground detection performance in video of our proposed method in the ICS with the other popular foreground detection methods: the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106]. The F-measure scores show that our foreground detection method achieves better foreground detection accuracy than the other four popular methods.

We finally evaluate our proposed method using video frames that contain temporarily stopped vehicles. In particular, the first image in Figure 3.9 shows one frame from an

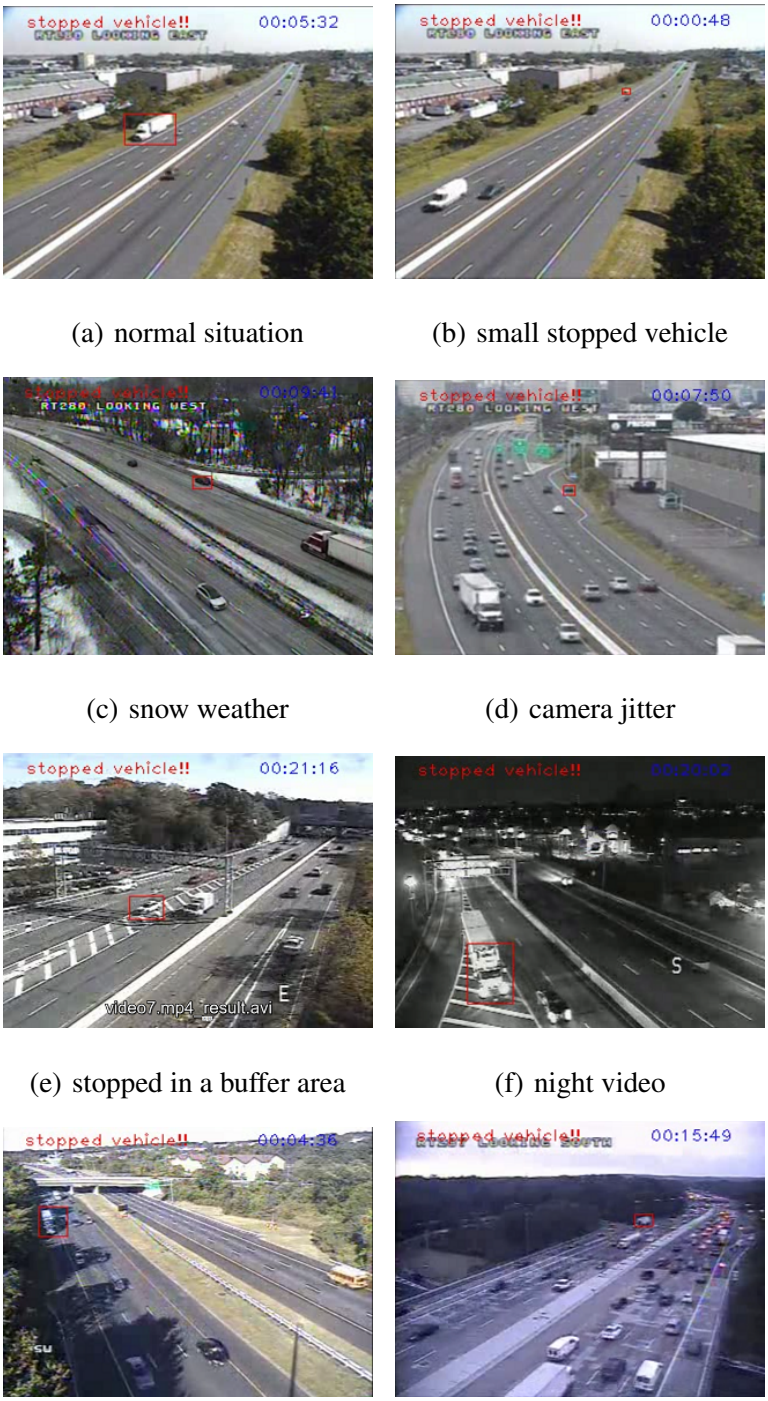


(a) (b)

(c) (d)

(e) (f)

Figure 3.9 Comparative performance of the proposed foreground detection method and other popular methods in a scene with some stopped vehicles. The First row shows one video frame from an NJDOT traffic video with spatial resolution of 704×480, and the detection result of our proposed method. The second row and the third row display the detection results using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106], respectively.



(a) normal situation (b) small stopped vehicle

(c) snow weather (d) camera jitter

(e) stopped in a buffer area (f) night video

(g) strong shadow (h) illumination change

Figure 3.10 Some stopped vehicle detection results. The stopped vehicles detected by our proposed method are marked with the red rectangles.

NJDOT traffic video with some stopped vehicles, and the second image on the first row in Figure 3.9 shows the foreground detection result of our method using the 12-dimensional features constructed from the ICS. The second row and the third row in Figure 3.9 show the detection results using the GMM method [107], the Zivkovic's method [136], [137], the Hayman and Eklundh's method [44], and the PAWCS method [106], respectively. Figure 3.9 shows that the four popular methods did not detect the stopped vehicles or could only detect a little part of the stopped vehicles. Our foreground detection method, however, is able to correctly detect those temporarily stopped vehicles.

The property of our proposed method for detecting stopped targets has broad applications in video analysis in general and in traffic incidents detection in particular. For example, the popular statistical modeling methods usually have difficulty in detecting traffic incidents such as congestion or stopped vehicles due to the lack of such property [121], [129]. Our proposed method, in contrast, is capable of detecting these traffic incidents. Specifically, we have implemented experiments using 30 videos of 30 minutes or 60 minutes from NJDOT for stopped vehicle detection. These video sequences contain different real traffic situations and video qualities, such as low video resolution, bad video quality, camera jitter, and bad weather conditions. Most of these real world videos do not have good quality, this causes most of the vehicle detection methods can not identify the vehicles. The detailed description of the video sequences is shown in Table. 3.5. Among these videos, the stopped vehicle incident occurred 22 times, ranging from 10 seconds to 15 mins. The long stopping time causes some other foreground detection methods cannot keep detecting the stopped vehicles, even they can detect some temporarily stopped foreground objects. In contrast, our proposed method is able to detect those stopped vehicles as long as they are stopping there. Our method is able to detect 21 out of 22 of these stopped vehicle incidents, and no false positive detection occurred. The only one we miss is because of the night vision and highly blurred video frames. We show some stopped vehicle detection results in Figure 3.10. These figures include several challenging conditions in real world

Table 3.5 The Description of the Video Sequences We used in Experiments

resolution	frame rate (fps)	duration (mins)	bit rate (kbps)	condition	count
320×240	15	30	45 to 132	normal	12
320×240	15	30	41 to 131	night time	4
320×240	15	30	47 to 176	strong shadow	9
320×240	15	30	123	fog	1
320×240	15	30	82	rain	1
640×480	15	30	1066	snow	1
352×480	30	60	633 - 839	normal	2

traffic videos, such as low resolution, bad weather conditions, camera jitter, night video, shadow, etc. We can see that our proposed method can detect the stopped vehicles under all these conditions without false positive detections.

3.4 Conclusions

We have presented in this chapter a novel foreground detection method for video analysis by integrating color, wavelet, and temporal features.

First, a local background modeling (LBM) process, which capitalizes on the traditional Gaussian mixture models, is explained by choosing the most significant single Gaussian density to model the background locally for each pixel according to the weights learned for the Gaussian mixture model. Then a novel Global Foreground Modeling (GFM) method is presented to model the foreground, which estimates a global probability density function for the foreground and applies the Bayes decision rule for minimum error to choose one Gaussian density function for a specific time and location. Additionally, the Bayes classifier is applied for the classification of foreground and background pixels.

Second, to mitigate the correlation effects of the RGB color space on the independence assumption among the color component images, the YIQ, the YCbCr, and some uncon-

ventional color spaces such as the Uncorrelated Color Space (UCS), the Independent Color Space (ICS), and the Discriminating Color Space (DCS) are investigated for feature extraction. Note that the YIQ and the YCbCr separate the chromatic and the achromatic values, the UCS and DCS further decorrelate the component images, and the ICS satisfies the independence assumption by deriving the independent component images.

Third, to further enhance the discriminatory power of the feature vectors, the horizontal and vertical Haar wavelet features and the temporal difference features are integrated into the color features to build a new 12-dimensional feature vector. The 12-dimensional feature vector thus is able to increase the discriminatory power, which helps our foreground detection method achieve better performance than other popular statistical modeling methods.

As a result, the proposed method is able to address the challenging problems, such as better satisfying the independence assumption in statistical modeling, insufficient discriminatory power of the input feature vector in the RGB color space, the inappropriate statistical modeling of the foreground, and the final classification of the foreground and background pixels.

Experimental results using videos from the NJDOT and a public video dataset show that the proposed foreground detection method improves upon the popular statistical modeling methods for foreground detection and is able to detect stopped moving objects.

CHAPTER 4

A NEW MOVING CAST SHADOW DETECTION METHOD FOR VIDEO ANALYSIS USING COLOR AND STATISTICAL MODELING

4.1 Introduction

In video analysis, shadows are often detected as part of the foreground, as they share similar motion patterns to the foreground objects [89], [94], [75]. These cast shadows often adversely affect the video analysis performance in various applications, such as tracking and object detection. Many algorithms have been published to detect the moving foreground objects in video [107], [44], [136], [103], [59], [128], [14], [13], [11], [98], [99]. Some methods like the Gaussian Mixture Modeling (GMM) estimate the background for each pixel using a number of Gaussian distributions [107], [44], [136], [103], [128], [14]. Other methods apply a classification method, such as the support vector machine (SVM), to classify the foreground and the background pixels [86], [128], [14]. Yet the cast shadows are usually classified into the foreground class as they have similar motion patterns to their foreground objects, which deteriorates video analysis performance.

In this chapter, we present a novel moving cast shadow detection method based on color and statistical modeling to detect and remove the cast shadows from the foreground region in order to improve video analysis performance. The novel moving cast shadow detection method contains four hierarchical steps, whose contributions are summarized below. First, we present a set of new chromatic criteria to detect the candidate shadow pixels in the HSV color space. We use the HSV color space for shadow detection due to its property of separating the chromaticity from intensity [29], [93], [21], [94], [38]. Our new chromatic criteria are more robust than the criteria used by other popular methods for shadow detection [94], [38]. Second, we present a new shadow region detection method to cluster the candidate shadow pixels into shadow regions. Many shadow detection methods can not solve the shadow outlines problem: the outlines of the shadow regions are often

classified to the foreground. As a result, after removing the shadow pixels from the foreground, the shadow regions are only partially removed, and the shadow outlines are often classified to the foreground. Our new shadow region detection method is able to solve this problem by applying the prior knowledge that both the foreground objects and their cast shadows should define continuous regions. Third, we present a statistical shadow modeling method, which uses a single Gaussian distribution to model the shadow class, to classify shadow pixels. The shadow pixels detected by both the new chromatic criteria and the new shadow region detection method tend to be more reliable shadow pixels, we therefore use these shadow pixels to estimate the Gaussian distribution for the shadow class. Finally, we present an aggregated shadow detection method that integrates the detection results using the new chromatic criteria, the new shadow region detection method, and the new statistical shadow modeling method. A gray scale shadow map is obtained by calculating a weighted summation of the candidate shadow pixels. A shadow free foreground may be derived by thresholding the gray scale shadow map.

We implement experiments using the public video data 'Highway-3' and the New Jersey Department of Transportation (NJDOT) traffic video sequences to show the feasibility of the proposed method. In particular, the experimental results (both qualitative and quantitative results) show that our proposed method achieves better shadow detection performance than some popular shadow detection methods [48], [62], [49], [95], [38].

4.2 A Novel Moving Cast Shadow Detection Method

In video analysis, shadows are often detected as part of the foreground, which deteriorates the performance of many video analysis tasks. We therefore present a novel moving cast shadow detection method that is able to detect and remove the cast shadows from the foreground.

In particular, Figure 4.1 shows the system architecture of our proposed moving cast shadow detection method in video using color and statistical modeling. First, we apply the

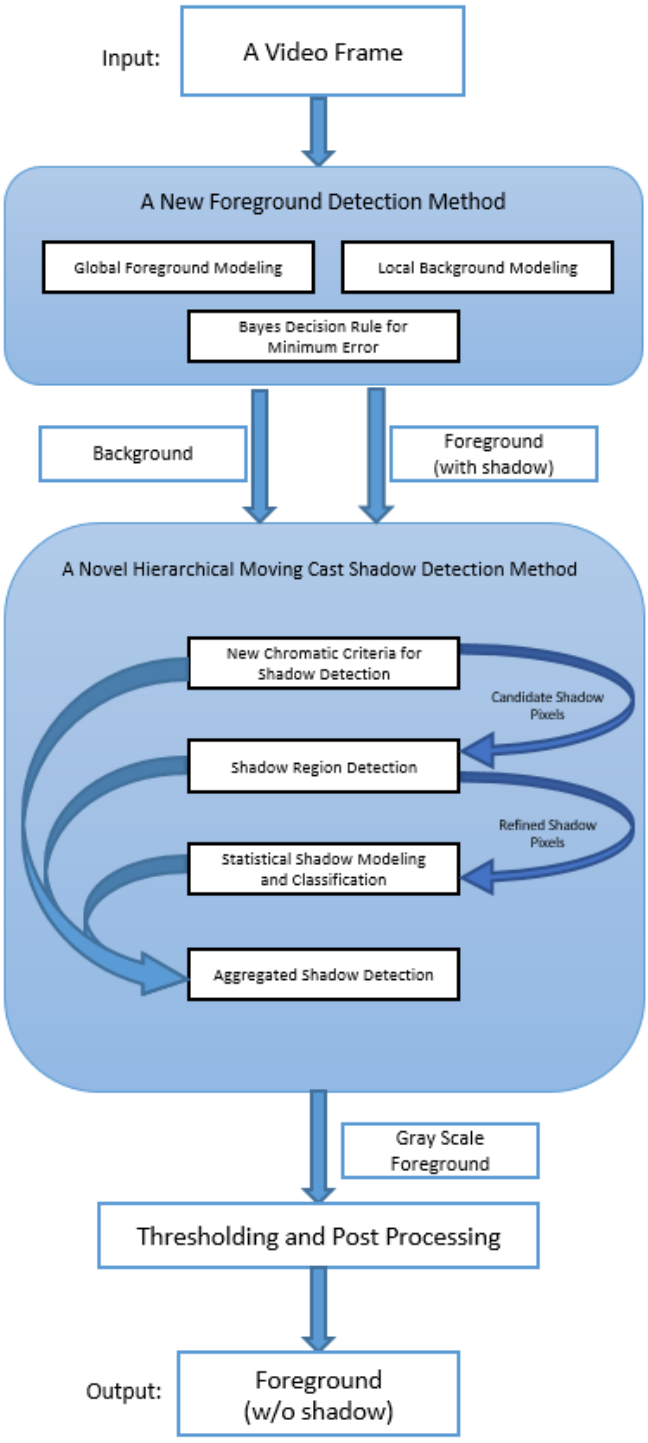
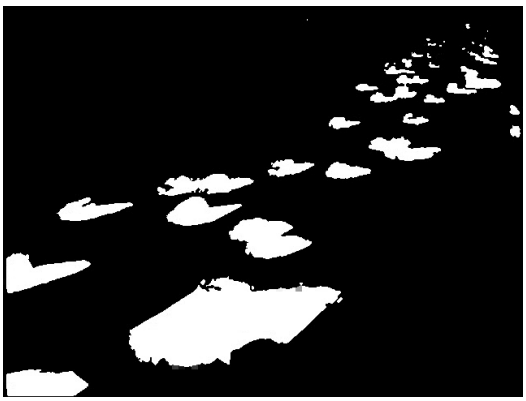


Figure 4.1 The system architecture of our novel moving cast shadow detection method.

(a)

(b)



(c)

Figure 4.2 (a) A video frame from an NJDOT traffic video. (b) The background derived using the GMM model. (c) The foreground (with shadow) detected using our new foreground detection method.

foreground detection method introduced in the previous chapter to detect the foreground, which contains both the foreground objects and their cast shadows. Figure 4.2 (a) shows a video frame from an NJDOT traffic video. The background, which is derived using the GMM model, is shown in Figure 4.2 (b), and the foreground (with shadow), which is detected using our new foreground detection method, is displayed in Figure 4.2 (c). As shown in Figure 4.2, the foreground includes both the foreground objects and their cast shadows. Second, we present a moving cast shadow detection method with the following novelties: (i) A new method based on new chromatic criteria is presented for candidate shadow pixel detection. (ii) A shadow region detection method is proposed to cluster the candidate shadow pixels into shadow regions. (iii) A statistical shadow model is presented for classifying shadow pixels. (iv) An aggregated shadow detection method is presented for final shadow detection.

4.2.1 New Chromatic Criteria for Shadow Pixel Detection

As color provides useful information for shadow detection, we present in this subsection a new method based on a set of new chromatic criteria for shadow pixel detection. After foreground detection, we need to detect the cast shadow pixels in the foreground region. Our new method will apply the new chromatic criteria to detect candidate shadow pixels. As the HSV color space is widely used in shadow detection due to its property of separating the chromaticity from intensity, we choose this color space for shadow detection.

As shown in Figure 4.3, the HSV color space can be modeled as a cone in geometry. Let H, S, and V be the H (hue), S (saturation), and V (value) components in the HSV color space. H represents color information, which is described by the angular dimension of the cone. S denotes the concentration of color, increasing from the central vertical axis to the edge of the cone. V represents the brightness of a pixel, from the darkest at the bottom to the brightest at the top.

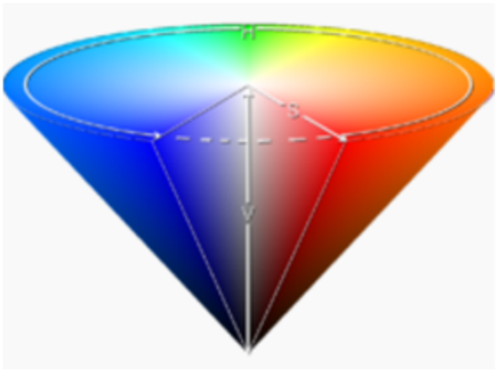


Figure 4.3 The HSV color space can be modeled as a cone.

Let S_f and V_f be the S and V components of a pixel in the foreground region, respectively, and S_b and V_b be the S and V components of the same pixel in the background, respectively. Our new chromatic criteria are defined as follows:

$$\begin{cases} \tau_{sl} < S_f - S_b < \tau_{sh} \\ \tau_{vl} < V_b - V_f < \tau_{vh} \end{cases} \tag{4.1}$$

where τ_{sl}, τ_{sh}, τ_{vl}, and τ_{vh} represent the thresholds. If a pixel in the foreground region satisfies these chromatic criteria, it is classified as a candidate shadow pixel.

To illustrate the rationale of our new chromatic criteria, we show the difference of the S component between the foreground and the background, and the difference of the V component between the background and the foreground, respectively. In particular, Figure 4.4 (a) shows a color video frame, Figure 4.4 (b)-(d) display the H (hue), S (saturation), and V (value) components in the HSV color space, Figure 4.4 (e) shows the difference of the S component between the foreground and the background, and Figure 4.4 (f) shows the difference of the V component between the background and the frame. From Figure 4.4 (e), we can see that for the shadow pixels the difference values of the S component between the frame and the background are within a range that can be bounded by two threshold values τ_{sl} and τ_{sh} as shown in Equation (4.1). From Figure 4.4 (f), we can see that for the shadow pixels the difference values of the V component between the

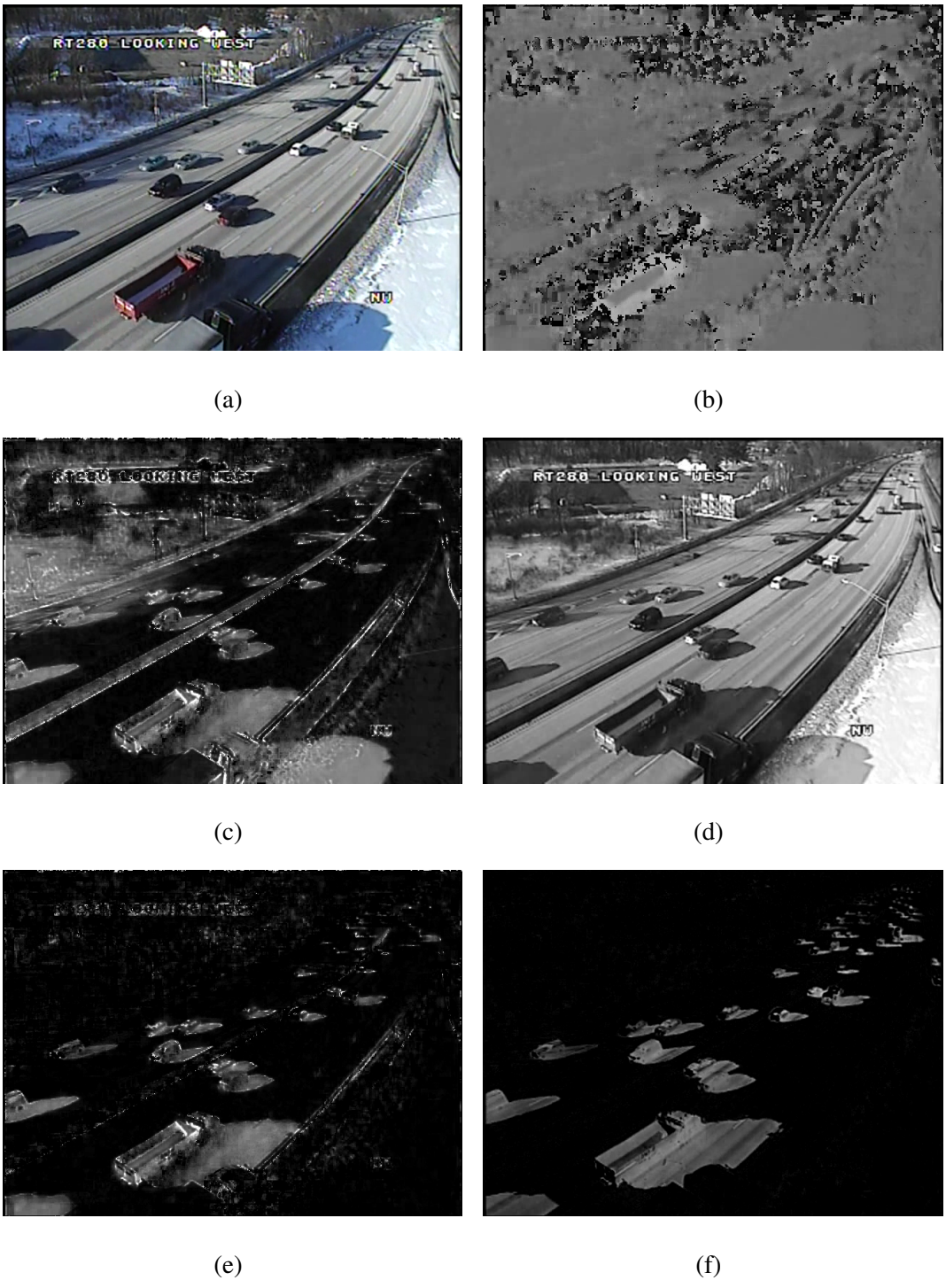


Figure 4.4 (a) The H (hue) component of the video frame. (b) The S (saturation) component of the video frame. (c) The V (value) component of the video frame. (d) The difference of the S component between the frame and the background. (e) The difference of the V component between the background and the frame.

background and the frame also fall into a range that can be bounded by two threshold values τ_{vl} and τ_{vh} as shown in Equation (4.1).

Note that many shadow detection methods assume that the shadow areas are darker in intensity but relatively invariant in chromaticity [29], [93], [21], [94], [38]. As a result, some color spaces that separate chromaticity from intensity are applied to detect shadows, such as the HSV color space [29], the c1c2c3 color space [93], and the YUV color space [21]. Some popular methods [94], [38] apply a different set of chromatic criteria: $|H_f - H_b| \leq \tau_H$, $S_f - S_b \leq \tau_S$, $\beta_1 \leq V_f/V_b \leq \beta_2$, where H_f, S_f, V_f, H_b, S_b and V_b represent the hue, saturation, and value of a pixel of the frame and the background, respectively. τ_H, τ_S, β_1 and β_2 are the thresholds that are chosen empirically. The pixels that satisfy these three criteria are classified as shadow pixels. These chromatic criteria assume that the cast shadows have a similar hue to the background, but a lower S (saturation) and a lower V (value) than the background [94].

In contrast, our new chromatic criteria are more robust than these chromatic criteria. In our research, we find that the assumption that the cast shadows have a similar hue to the background is often not satisfied. For example, Figure 4.4 (b) shows that the H values of the cast shadows are not similar to the background. As a result, in our new chromatic criteria, the H values are excluded as they vary a lot, especially for the background. The S values, however, are relatively stable for the background and the cast shadows comparatively, but vary for the foreground objects. Thus the difference of the S component between the shadow and the background often falls into a fixed range. Another characteristic of cast shadows is that the shadows are always darker than the background, but they cannot be exactly black. Based on these observations, we present our new chromatic criteria for candidate shadow pixel detection as shown in Equation (4.1).

Figure 4.5 shows the shadow detection results using our new chromatic criteria and the criteria in [94], [38]. Specifically, Figure 4.5 (a) shows a video frame from an NJDOT traffic video, Figure 4.5 (b) displays the shadow detection results using the chromatic

(a)

(b)

(c)

Figure 4.5 (a) A video frame from an NJDOT traffic video. (b) The shadow detection results (shadow pixels are represented using gray scale value of 128) using the chromatic criteria in [94], [38]. (c) The shadow detection results using our new chromatic criteria.

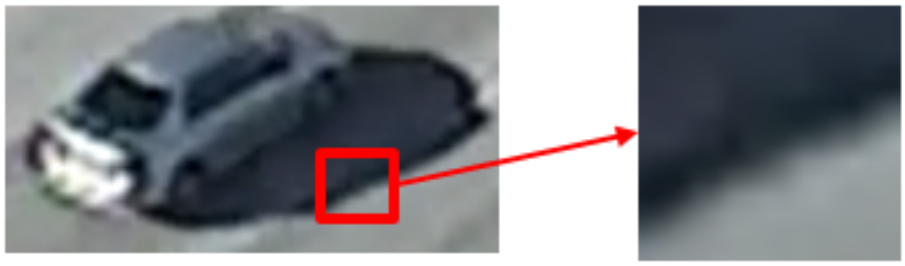


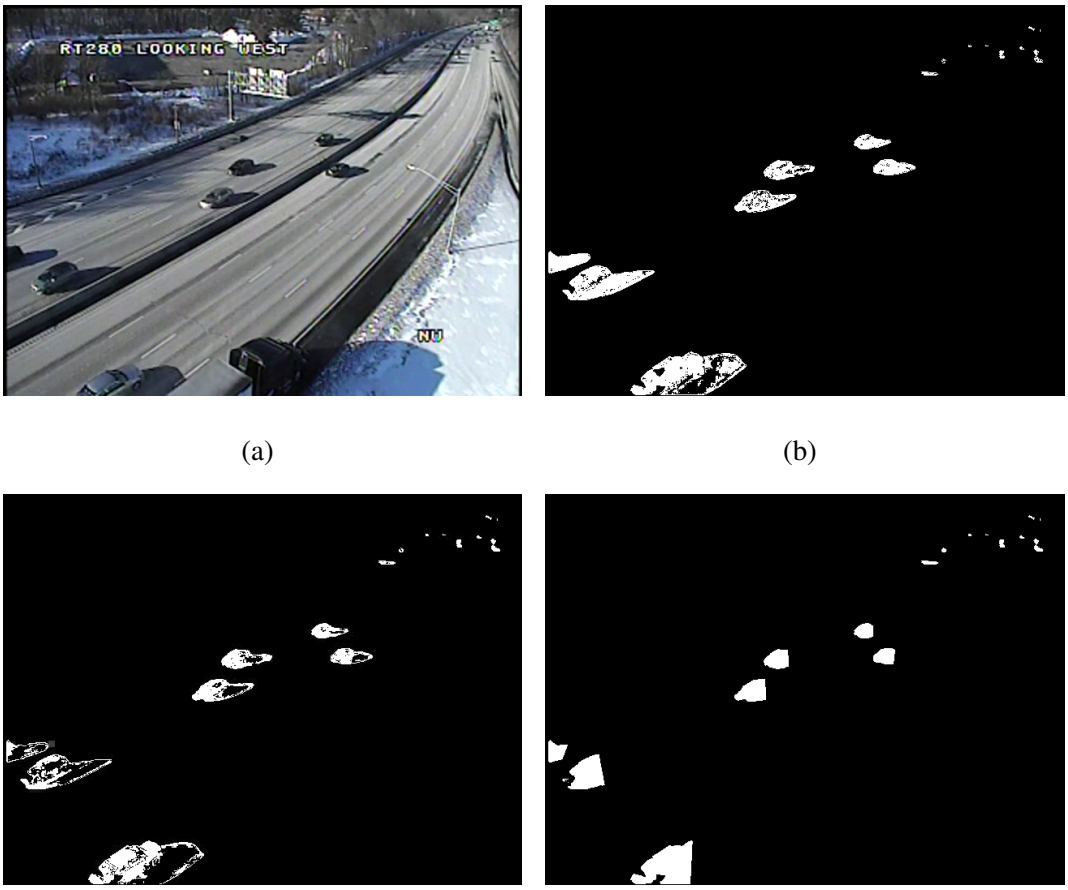
Figure 4.6 The color of the outline is different from that of the main part of the shadow.

criteria in [94], [38], and Figure 4.5 (c) shows the shadow detection results using our new chromatic criteria. Note that the shadow pixels are represented using gray scale value of 128. We can see from Figure 4.5 (b) and (c) that our proposed method using the new chromatic criteria is able to detect the shadow pixels more reliably.

4.2.2 A New Shadow Region Detection Method

One inherent problem in shadow detection is that the outlines of the shadow region are often classified to the foreground class. We can see from Figure 4.6, the color of the outline is different from that of the main part of the shadow. As a result, after removing the shadow pixels from the foreground, the shadow regions are only partially removed, and the shadow outlines are often classified to the foreground. Figure 4.7 (b) and (c) show the partially removed shadow regions and the shadow outlines that are not removed. These unremoved shadow regions and outlines often deteriorate the performance of video analysis tasks, such as video tracking and incident detection.

To solve this problem, we present a new shadow region detection method based on the prior knowledge that both the foreground objects and their cast shadows should define continuous regions. Note that in each frame, the detected foreground often consists of several foreground regions, each of which contains both the foreground objects and their



(a) (b)

(c) (d)

Figure 4.7 (a) A video frame from an NJDOT traffic video. (b) The shadow detection results using Huang and Chen's method [49]. (c) The shadow detection results using our new chromatic criteria. (d) The shadow detection results using our shadow region detection method..

cast shadows. In each foreground region, all the shadow pixels are on one side and all the foreground object pixels are on the other side. As a result, each foreground region may be divided into two regions: the shadow region and the foreground object region. As the candidate shadow pixels inside each foreground region are detected using the new chromatic criteria introduced in subsection 4.2.1, the remaining pixels are the foreground object pixels.

The idea of our new shadow region detection method is to cluster the shadow pixels and the foreground object pixels into two classes using the centroids of the two classes.

Our idea is similar to the K-means clustering algorithm but without any iteration steps. Specifically, in each foreground region B, we first find the centroid of the candidate shadow pixels $Cent_S(B)$ and the centroid of the foreground pixels $Cent_O(B)$. We then compute the Euclidean distances between each pixel and the two centroids. We finally classify the pixel into a foreground object class or a shadow class based on the Euclidean distances: if the distance to the foreground object class is smaller, the pixel is assigned to the foreground object class, and vice versa. In particular, for the pixel $\mathbf{x}$ at location (i, j) in each foreground region B, we calculate the distance between the pixel and the shadow centroid $Dist(\mathbf{x}_{ij}, Cent_S(B))$ and the distance between the pixel and the foreground object centroid $Dist(\mathbf{x}_{ij}, Cent_O(B))$, respectively. If $Dist(\mathbf{x}_{ij}, Cent_S(B))$ is smaller, than we classify $\mathbf{x}_{ij}$ into the shadow class. Otherwise, we classify it into the foreground object class. The new shadow region detection method thus detects the candidate shadow regions.

Figure 4.7 (a) displays a video frame from an NJDOT traffic video, Figure 4.7 (b) shows the shadow detection results using Huang and Chen's method [49], Figure 4.7 (c) shows the shadow detection results using the new chromatic criteria introduced in subsection 4.2.1, and Figure 4.7 (d) shows the shadow detection results using the new shadow region detection method. Figure 4.7 (b) and (c) reveal that the outlines of the shadow region are often classified to the foreground class, which leads to an incorrect shadow detection. In contrast, Figure 4.7 (d) shows that our proposed new shadow region detection method is able to detect the whole shadow regions including their outlines.

4.2.3 A New Statistical Shadow Modeling and Classification Method

We present in this subsection a new statistical shadow modeling and classification method. For statistical modeling, we use a single Gaussian distribution to model the shadow class. In the previous two subsections, our proposed method using the new chromatic criteria detects candidate shadow pixels and our new shadow region detection method detects the candidate shadow regions. As the shadow pixels detected in both methods tend to be more

reliable shadow pixels, we apply these shadow pixels to estimate the Gaussian distribution for the shadow class.

Specifically, let $\mathbb{S}_c$ and $\mathbb{S}_r$ be the candidate shadow pixel sets detected by our proposed method using the new chromatic criteria and our new shadow region detection method, respectively. For each pixel $\mathbf{x}$ in the foreground, if $\mathbf{x} \in \mathbb{S}_c$ and $\mathbf{x} \in \mathbb{S}_r$, we will use $\mathbf{x}$ to update the Gaussian distribution $N_s(\mathbf{M}, \Sigma)$ as follows:

$$\mathbf{M}' = \mathbf{M} - \alpha(\mathbf{M} - \mathbf{x}) \tag{4.2}$$

$$\Sigma' = \Sigma + \alpha((\mathbf{M} - \mathbf{x})(\mathbf{M} - \mathbf{x})^t - \Sigma) \tag{4.3}$$

where $\mathbf{M}$ and Σ are the mean vector and the covariance matrix of the shadow Gaussian distribution, respectively. α is a number between 0 and 1 which influences the model updating speed.

For shadow pixel classification, we apply the following discriminant function for each pixel $\mathbf{x} \in \mathbb{R}^d$ in the foreground:

$$s(v_i) = (\mu_i - v_i)^2 - p\,\sigma_i^2 \qquad i \in \{1, 2, \ldots, d\} \tag{4.4}$$

where v_i is the i-th element of the input vector $\mathbf{x}$, μ_i is the i-th element of the mean vector $\mathbf{M}$, σ_i^2 is the i-th diagonal element of the covariance matrix Σ, and p is the parameter which determines the threshold. If $s(v_i)$ is greater than zero for any $i \in \{1, 2, \ldots, d\}$, we classify $\mathbf{x}$ into the foreground object class. Otherwise we classify it as a shadow pixel. Our new statistical shadow modeling and classification method thus detects the candidate shadow pixels.

4.2.4 Aggregated Shadow Detection

The final step for cast shadow detection is the aggregated shadow detection that integrates the detection results using the new chromatic criteria, the new shadow region detection method, and the new statistical shadow modeling and classification method discussed in the previous three sections. Specifically, we first assign all the pixels in the shadow class a gray scale value of 128, and the pixels in the foreground object class a gray scale value of 255. We then define three weights for the three methods to indicate their significance for the final cast shadow detection: w_c for the new chromatic criteria, w_r for the shadow region detection, and w_s for the statistical modeling. The weights are normalized so that their summation equals one:

$$w_c + w_r + w_s = 1 \tag{4.5}$$

Note that the larger a weight is, the greater impact the corresponding method exerts to the final shadow detection results. These weights may be learned from the data, but without any prior information, they may be set to equal values.

For each location (i, j) in the foreground, the gray level $G(i, j)$ is calculated as follows:

$$G(i, j) = w_c C(i, j) + w_r R(i, j) + w_s S(i, j) \tag{4.6}$$

where $C(i, j)$, $R(i, j)$ and $S(i, j)$ are the values at location (i, j) derived by using the new chromatic criteria, the shadow region detection, and the statistical modeling method, respectively.

In the gray-scale image, the smaller value a pixel has, the more likely it is a shadow pixel. We use a threshold T_s to generate a shadow free binary foreground mask. T_s can be determined empirically, and it can also be obtained by training some training samples. The

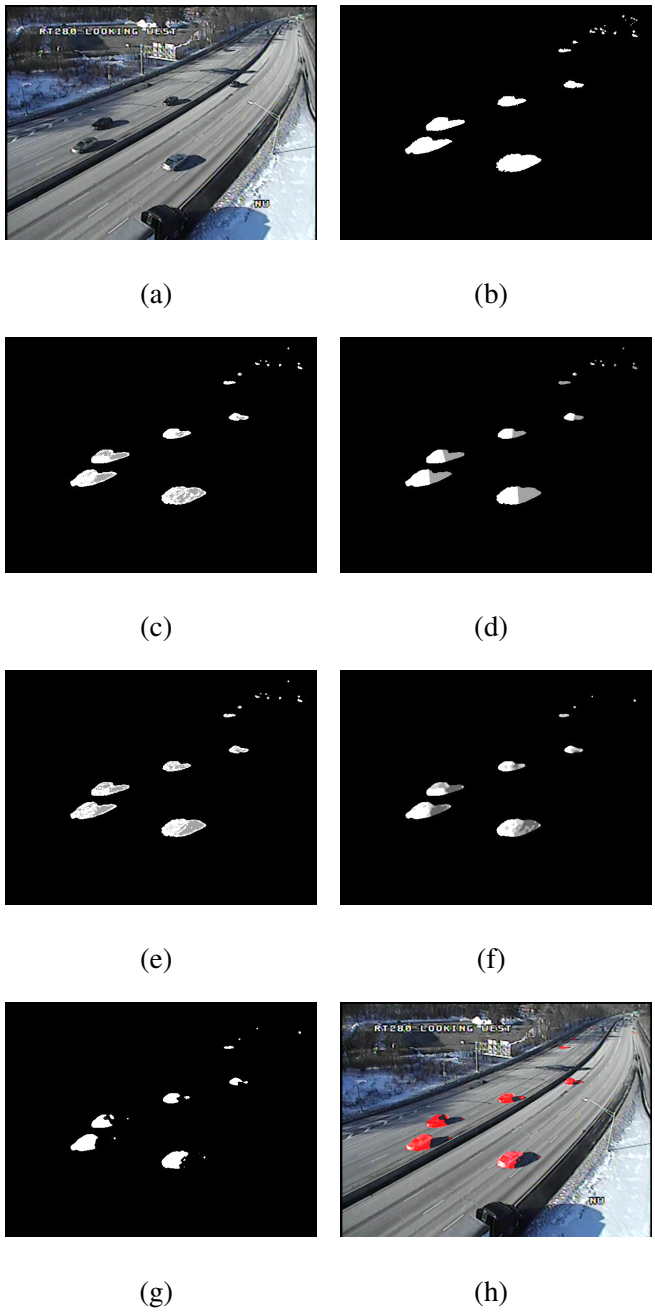


Figure 4.8 (a) A video frame from the NJDOT traffic video. (b) The detected foreground (with shadow) using the new foreground detection method. (c) The detected shadow pixels using the new chromatic criteria. (d) The detected shadow regions using the shadow region detection method. (e) The detected shadow pixels using statistical shadow modeling and classification. (f) The detected shadow pixels using the aggregated shadow detection method. (g) The shadow free foreground. (h) The video frame that shows foreground pixels in red color.

binary value $B(i,j)$ at location (i,j) is calculated as follows:

$$B(i,j) = \begin{cases} 0, & if\ G(i,j) < T_s \\ 255, & otherwise \end{cases} \tag{4.7}$$

Figure 4.8 shows the results of our novel cast shadow detection method step by step. Figure 4.8 (a) is a video frame from an NJDOT traffic video. Figure 4.8 (b) shows the foreground detected using our new foreground detection method. Figure 4.8 (c)-(e) show the shadow detection results using the chromatic criteria detection, the shadow regions detection, and the statistical modeling detection. Note that the shadow pixels are indicated using the gray scale value of 128. Figure 4.8 (f) shows the gray-scale image generated by the aggregated shadow detection method. Figure 4.8 (g) shows the foreground after removing the shadows. Figure 4.8 (h) displays the video frame with the foreground in red color.

4.3 Experiments

We first show the quantitative evaluation results using a challenging video, the 'Highway-3' video [94]. This video, which is publicly available and broadly used, facilitates the comparative evaluation of our proposed method with other representative shadow detection methods published in the literature. We then use the New Jersey Department of Transportation (NJDOT) traffic video sequences to evaluate our proposed method qualitatively. Specifically, we apply four NJDOT traffic videos, each of which is 15 minutes with a frame rate of 15 frames per second or fps. The computer we use is a DELL XPS 8900 PC with a 3.4 GHz processor and 16 GB RAM. The 'Highway-3' video has a spatial resolution of 320×240, and it takes 9ms to process each frame using our method. The NJDOT videos have a spatial resolution of 640×482, and it takes 39ms to process each frame using our method. As a result, our proposed shadow detection method is able to perform real time analysis of these videos.

Table 4.1 The Comparative Shadow Detection Performance of Our Proposed Method and Some Popular Shadow Detection Methods

Methods	η	ξ	$F-measure$
Bullkich *et al.* [18]	80%	61%	69%
Lalonde *et al.* [58]	39%	86%	54%
Guo *et al.* [41]	42%	82%	55%
Sanin *et al.* [95]	62%	**91%**	74%
Gomes *et al.* [38]	65%	90%	75%
Our Proposed Method	**90%**	76%	**83%**

The shadow detection rate η, the shadow discrimination rate ξ, and the F-measure are popular metrics used to evaluate shadow detection performance quantitatively [52], which are defined as follows:

$$\eta = \frac{TP_s}{TP_s + FN_s} \tag{4.8}$$

$$\xi = \frac{TP_o}{TP_o + FN_o} \tag{4.9}$$

$$F - measue = \frac{2\eta\xi}{\eta + \xi} \tag{4.10}$$

where TP_s and FN_s represent the number of true positive and false negative shadow pixels, respectively, and TP_o and FN_o stand for the number of true positive and false negative object pixels, respectively.

Table. 4.1 shows the comparative shadow detection performance of our proposed method and some popular shadow detection methods using the publicly available challenging 'Highway-3' video. In particular, our proposed method achieves the highest shadow detection rate of 90%, compared with the 80%, 39%, 42%, 62%, and 65% shadow detection rates by the Bullkich *et al.* [18] shadow detection method, the Lalonde *et al.* [58] shadow detection method, the Guo *et al.* [41] shadow detection method, the Sanin *et al.* [95] shadow

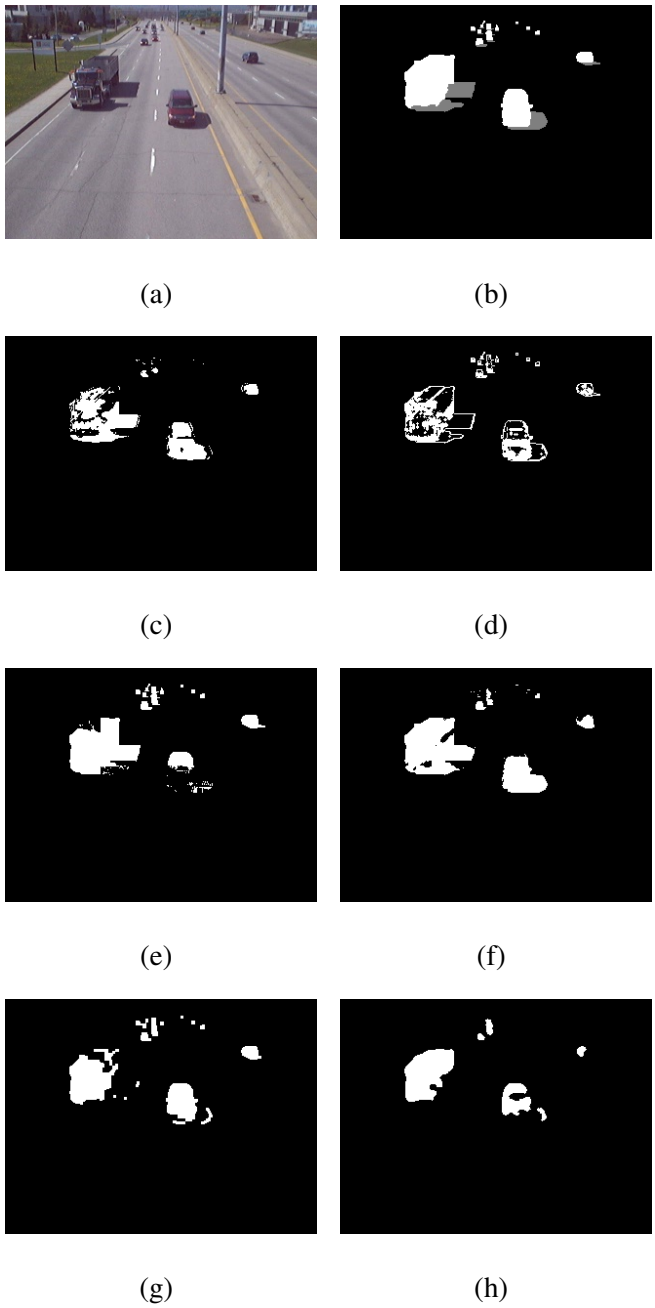


Figure 4.9 The foreground masks obtained by different methods. (a). One video frame of 'Highway-3' video [94]. (b). The ground truth of the foreground mask. The white parts are the foreground objects. The gray parts are the cast shadows. (c)-(g) The shadow free foreground mask of Cucchiara *et al.*'s method [29], Huang and Chen's method [49], Hsieh *et al.*'s method [48], Leone and Distante's method [62], and Sanin *et al.*'s method [95], respectively. (h)The shadow free foreground mask of our proposed method.

detection method, and the Gomes *et al.* [38] shadow detection method, respectively. Our proposed method also achieves the highest F-measure score of 83%, compared with the 69%, 54%, 55%, 74%, 75% F-measure scores by the Bullkich *et al.* [18] shadow detection method, the Lalonde *et al.* [58] shadow detection method, the Guo *et al.* [41] shadow detection method, the Sanin *et al.* [95] shadow detection method, and the Gomes *et al.* [38] shadow detection method, respectively.

Figure 4.9 shows the experimental results using a frame from the 'Highway-3' video. Specifically, Figure 4.9 (a) shows a video frame the 'Highway-3' video [94]. Figure 4.9 (b) shows the ground truth of the foreground mask, where the white regions represent the foreground objects and the gray regions represent the cast shadows. Figure 4.9 (c) shows the shadow free foreground mask by using Cucchiara *et al.*'s method [29]. Figure 4.9 (d) shows the shadow free foreground mask by using Huang and Chen's method [49]. Figure 4.9 (e) shows the shadow free foreground mask by using Hsieh *et al.*'s method [48]. Figure 4.9 (f) shows the shadow free foreground mask by using Leone and Distante's method [62]. Figure 4.9 (g) shows the shadow free foreground mask by using Sanin *et al.*'s method [95]. Figure 4.9 (h) shows the shadow free foreground mask by using our proposed shadow detection method. We can see from Figure 4.9 that our proposed shadow detection method achieves better shadow detection and removal results than the other popular shadow detection methods.

Another dataset we apply in our experiments is the NJDOT traffic video sequences. The videos in this dataset have stronger cast shadows and lower video quality than the 'Highway-3' video. Many shadow detection methods fail to detect shadows in these videos, but our proposed method is able to achieve good shadow detection performance on these videos as shown in Figure 4.8. The significance of shadow detection in these videos is to improve the performance of video analysis tasks such as tracking and object detection. In particular, Figure 4.10 shows comparatively the vehicle tracking performance using the NJDOT traffic videos: the vehicle tracking results without shadow detection and the vehicle

(a)

(b)

Figure 4.10 The comparison of vehicle tracking performance using a frame from the NJDOT traffic videos. (a) The vehicle tracking results without shadow detection. (b) The vehicle tracking results with shadow detection using our proposed shadow detection method.

tracking results with shadow detection using our proposed shadow detection method. We can see in the Figure 4.10 (a) that two vehicles are connected together by their cast shadows and fall into one tracking block when no shadow detection algorithm is applied. After applying our shadow detection algorithm, these two vehicles are separated into two tracking blocks. As a result, the tracking performance is more accurate.

4.4 Conclusions

We have presented in this chapter a novel moving cast shadow detection method for video analysis using new chromatic criteria and statistical modeling. The major contributions of our proposed method are four-fold.

First, we propose a set of new chromatic criteria for shadow pixels differentiation. Second, we use a shadow region detection method to detect the continuous shadow regions based on the property of cast shadows. Third, we build a statistical shadow model to model and classify the shadow pixels with a single Gaussian distribution. The model keeps learning and updating to adapt to the changes of the environment. Fourth, we use an aggregated shadow detection method to combine the shadow detection results from the previous three steps. A weighted summation strategy is used to aggregate the candidate shadow detection results.

The experimental results using the publicly available 'Highway-3' video and the NJDOT video sequences have shown that (i) our proposed method achieves better shadow detection performance than other popular shadow detection methods, and (ii) our proposed method is able to detect cast shadows in low quality videos, such as the NJDOT videos, while in comparison other methods fail to detect the shadows.

CHAPTER 5

A STATISTICAL MODELING METHOD FOR ROAD RECOGNITION IN TRAFFIC VIDEO ANALYTICS

5.1 Introduction

Region of interest (RoI) is a widely used concept in video analysis. A region of interest is defined by a subspace of the entire video frame, which includes the part that people are most concerned about. To better recognize the activities in traffic, people always want to apply the video analysis algorithms within the RoI instead of the whole frame to reduce the computation complexity of the video analysis tasks [16]. Therefore, inter-cognitive communication [7], [8] between the human and artificial video analysis systems exists during the RoI selection. With the development of artificial intelligence and cognitive info-communications, an intra-cognitive communication mode [7], [8] becomes a more popular way in RoI detection. The RoI detection system recognizes the RoI and transfers the RoI information to the video analysis system. The communication between the two artificial cognitive systems can largely reduce the unbalanced cognitive capabilities between human beings and video analysis systems.

In traffic surveillance videos, the most widely used RoI is the region of the road. Because most of the traffic activities have happened in the road area, such as traffic congestion, wrong way vehicles, and traffic accidents. Manually selecting the RoI is a common approach when analyzing traffic surveillance videos. However, people need to select the road region as the RoI for every different camera location, and redefine the RoI when the camera changes the viewing angle. In order to reduce the manual work for selecting the RoI, many automatic road recognition methods have been proposed [79]. Some methods try to use vehicle motion information to segment the frame into active and inactive traffic regions. They estimate the road region by generating a map for active traffic regions based on the trajectories or foreground masks of the moving vehicles. This kind

of approach requires a sufficient number of vehicles to pass along the road. Therefore, the initial time required to gather the required information can vary based on the traffic flow. Some approaches use single images and try to fit linear or polynomial equations to the straight or curvy road boundaries and lane marks. These methods perform better in the case of in-vehicle cameras where the vanishing point is easier to estimate and they are limited to well-structured roads with visible and distinguishable sign-lines which is not always the case. Some methods try to estimate the road boundaries by extracting low-level image features (e.g. color, edge, texture). These methods are usually based on single images and low-level features analysis to classify the pixels or groups of pixels into road regions and non-road regions. They do not consider the structure or boundaries of the road and only tend to estimate the road area based on the color ([109], [2]), edge ([135], [2], [105]) and texture([96], [133]) of the road surface. Neural networks are also used for road recognition [82], [74]. This kind of supervised learning method requires numerous labeled data for the training process, which is hard to achieve.

In this chapter, we propose a statistical modeling method to recognize the road regions automatically in the traffic video analysis. First, we introduce a temporal feature guided statistical modeling method to build the road model. We use the temporal features in videos as guidance to automatically extract some sample data from the estimated background image. This sample data set mainly contains the features of the road, but also contains some other features. We build a Gaussian mixture model using this data set and further prune the model to get a statistical model for the road. Second, we propose a road recognition method, which can detect the road regions in a video frame. The detected road regions can be used as the RoI for traffic video analysis tasks. In the end, we use some real traffic video sequences from the New Jersey Department of Transportation (NJDOT) to evaluate the performance of our proposed method. The experimental results show that our method is able to detect the road regions accurately and robustly.

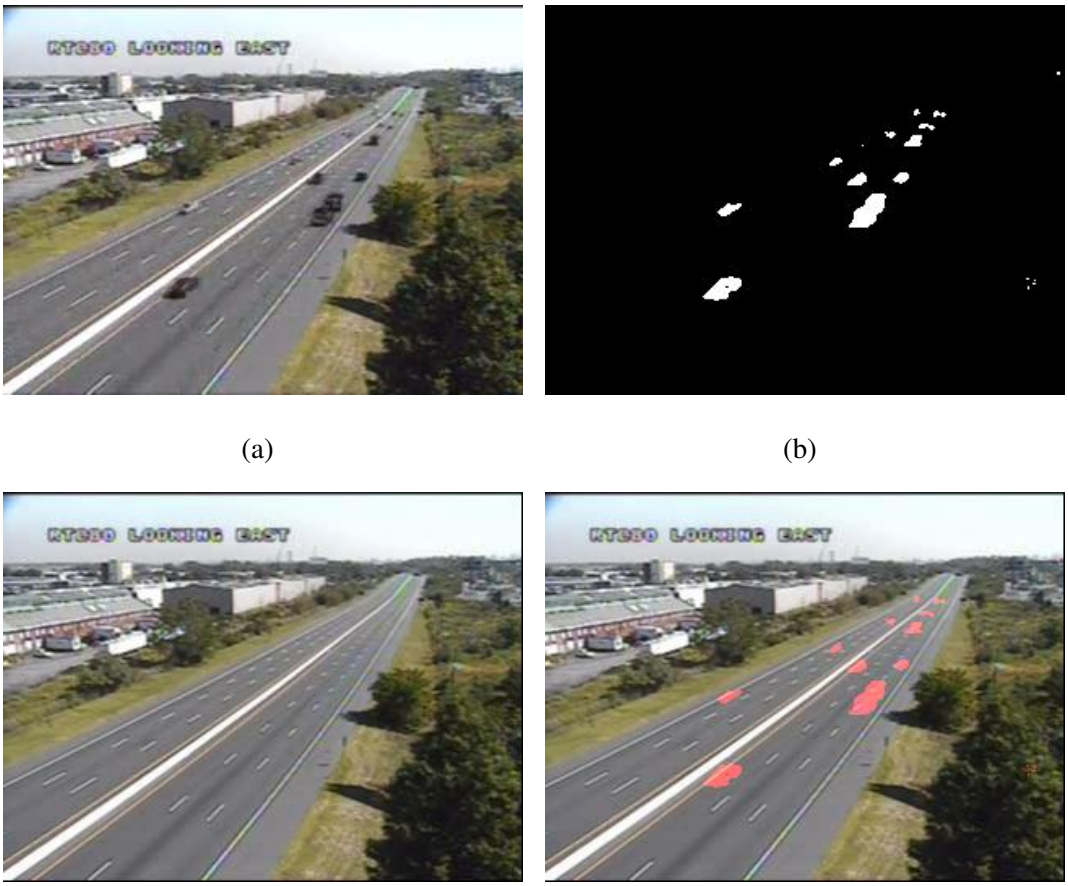
5.2 A Temporal Feature Guided Statistical Modeling Method for Road Recognition

The RoI selection is a widely used pre-processing technique of many video analysis methods. Manually selecting the RoI is a complex and tiresome task for human beings. Therefore, we propose a statistical modeling method for road recognition, which can detect the region of road automatically without any manual intervention. Our proposed method mainly has two major contributions: (i) The new temporal feature guided statistical modeling method can build the model without any label, which can reduce numerous manual work. (ii) The novel road recognition method can automatically segment the road region as the RoI for traffic video analysis.

5.2.1 The New Road Model Estimation Method

When building the statistical model for a class of objects, one common approach is to use some training data that has been labeled as this class to estimate the probability density function. This labeling work may require numerous efforts and time. Instead of using manually labeled data, we propose a temporal feature guided model estimation method, which can extract a sample data set from the video based on the temporal features.

In traffic videos, the region we are interested in is the road, which always has moving objects on it. One important information the moving objects can provide is temporal information. In order to utilize the temporal feature, we apply a foreground object detection method to segment the moving foreground objects and estimate the static background [99], [98]. The foreground detection method is able to detect the areas where have moving objects, and estimate a static background image that does not contain the moving objects. Figure 5.1 (a) shows a video frame from an NJDOT traffic video. Figure 5.1 (b) displays a binary foreground mask, where white pixels represent the moving foreground objects. Figure 5.1 (c) shows the estimated background image. Figure 5.1 (d) shows the corresponding regions of the moving foreground objects projected on the background image in red color.



(a) (b)

(c) (d)

Figure 5.1 (a) shows a video frame from an NJDOT traffic video. (b) displays a binary foreground mask, where white pixels represent the moving foreground objects. (c) shows the estimated background image. (d) shows the corresponding regions of the moving foreground objects projected on the background image in red color.

By projecting the moving foreground mask on the background image, we can get some regions, which contain the temporal features in the original video frame. As shown in Figure 5.1 (d), the red regions represent the projection of the foreground mask. We can see that most of these areas are road regions. Therefore, we can extract the feature vectors in these regions from the background image to build the road model.

For each video, we use the first N frames to build the model. Suppose $\mathbf{X} = \{\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}\}$ are the feature vectors we extracted from the N frames. As we only extract features from the regions corresponding to the foreground mask, $\mathbf{X}$ mainly contains the

features of the road at different locations and different times. However, some other features are still included in $\mathbf{X}$ due to the noises of the binary foreground mask, or overlapping caused by the viewing angle of the camera. We can use a Gaussian mixture model to estimate the distribution of sample set $\mathbf{X}$ as follows:

$$p\left(\vec{x}\right) = \sum_{m=1}^{M} \alpha_m \mathcal{N}\left(\vec{x}; \vec{\mu}_m, \Sigma_m\right) \tag{5.1}$$

where M is the number of components in the Gaussian mixture model, $\mathcal{N}\left(\vec{x}; \vec{\mu}_1, \Sigma_1\right), \ldots,$ $\mathcal{N}\left(\vec{x}; \vec{\mu}_M, \Sigma_M\right)$ are the Gaussian components. α_m is the weight of the m_{th} Gaussian components, and the summation of $\alpha_1, \ldots, \alpha_M$ is one. The Gaussian components are sorted in descending order according to the value of α.

Because the foreground detection result is not 100% accurate, the sample data we used to build the model is noisy. Some non-road features may also be involved in the sample data set. Therefore, we need to prune the Gaussian mixture model in order to get the road model. As we know, the majority of the sample set $\mathbf{X}$ is the feature of the road. The probability of the road features in the sample set is much higher than that of the noises. Therefore, the Gaussians with large weights can be used to describe the road. We select to use the first K Gaussians in $p\left(\vec{x}\right)$ as the road model, which is defined as follows:

$$p\left(\vec{x}|Road\right) = \frac{\sum_{k=1}^{K} \alpha_k \mathcal{N}\left(\vec{x}; \vec{\mu}_k, \Sigma_k\right)}{\sum_{k=1}^{K} \alpha_k} \tag{5.2}$$

K is defined as:

$$K = \arg\min_{k} \left(\sum_{m=1}^{k} \alpha_m > \left(1 - \mathcal{T}\right) \right) \tag{5.3}$$

where $\mathcal{T}$ is a threshold depends on the portion of the non-road features in the sample set $\mathbf{X}$. For example, if the foreground mask is noisy, we should select a high $\mathcal{T}$ value.

5.2.2 The Novel Road Recognition Method

In this subsection, we introduce a novel road recognition method using the statistical road model. As we know, the feature of the road is relatively simple. Most of the road regions look similar. The road model is based on the Gaussian mixture model, which has peaks at several feature points with the highest probability. If the feature vector of a pixel is close to any of these peaks, it would have a higher probability to be a road pixel. Otherwise, it is not a road pixel. By using this property of the road model, we propose a discriminant function to classify the feature vector $\vec{x}$ of each pixel into a road class and a non-road class. Suppose the road model contains K Gaussian distributions. The discriminant function is defined as follows:

$$R(\vec{x}) = \begin{cases} Road, & if \ \sum_{k=1}^{K} C(\vec{x})_k > 0 \\ Non-road, & otherwise \end{cases} \tag{5.4}$$

$$C(\vec{x})_k = \begin{cases} 1, & if \ D(\vec{x})_k < 0 \\ 0, & otherwise \end{cases} \tag{5.5}$$

$$D(\vec{x})_k = \sum_{i=1}^{d} (\vec{x_i} - \vec{\mu}_{k,i})^2 - \sigma_{k,i}^2 \tag{5.6}$$

where d represents the dimensionality of the feature vector,$\sigma_{k,i}$ means the i-th diagonal element of the covariance matrix Σ_k, $k \in \{1, 2, \ldots, K\}$.

We apply this discriminant function to every pixel in the estimated background image. We can build a road mask and assign every pixel classified as the road with 255, all the other pixels with 0. Then we can get a binary road mask showing the road pixels.

As we know, the road region is always a large continuous region in traffic surveillance videos. However, some pixels on the road may have abnormal features that can not be described by the road model, such as damaged areas, shadows, lane marks, etc. The miss classification of these pixels may cause the road mask to have some holes in the road region.

In addition, some non-road pixels may be detected as the road because they have similar features as the road. This will cause some noises outside the road region. In order to solve these issues, we apply a morphological operator on the road mask to further enhance the road recognition result. The morphological operation is defined as follows:

$$R'_{mask} = (R_{mask} \ominus E) \oplus D \tag{5.7}$$

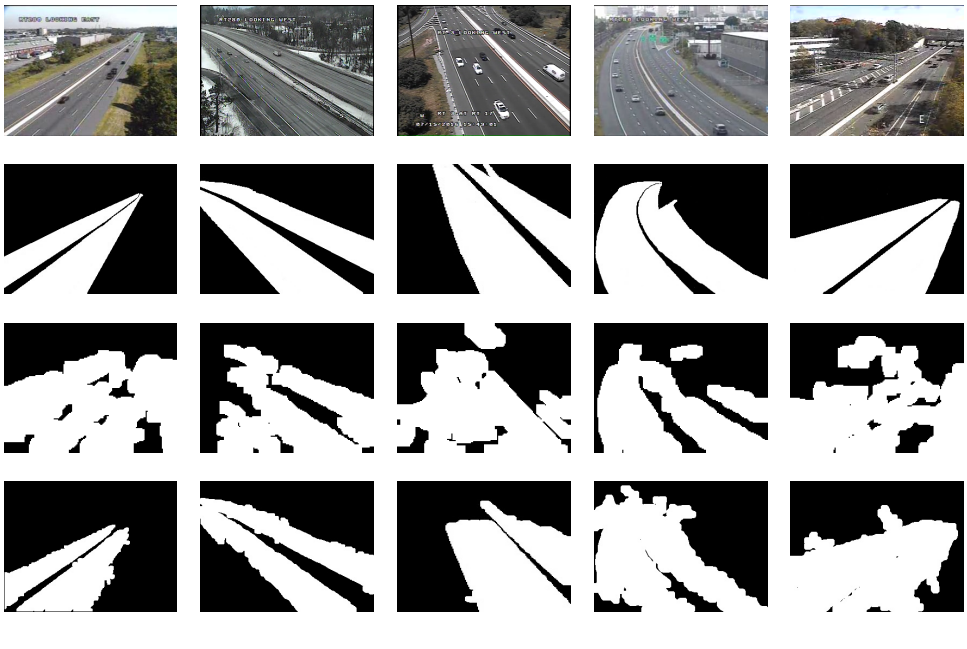
where R_{mask} is the road mask, $\ominus$ is the erosion operator, E is the erosion template, $\oplus$ is the dilation operator, and D is the dilation template.

5.3 Experiments

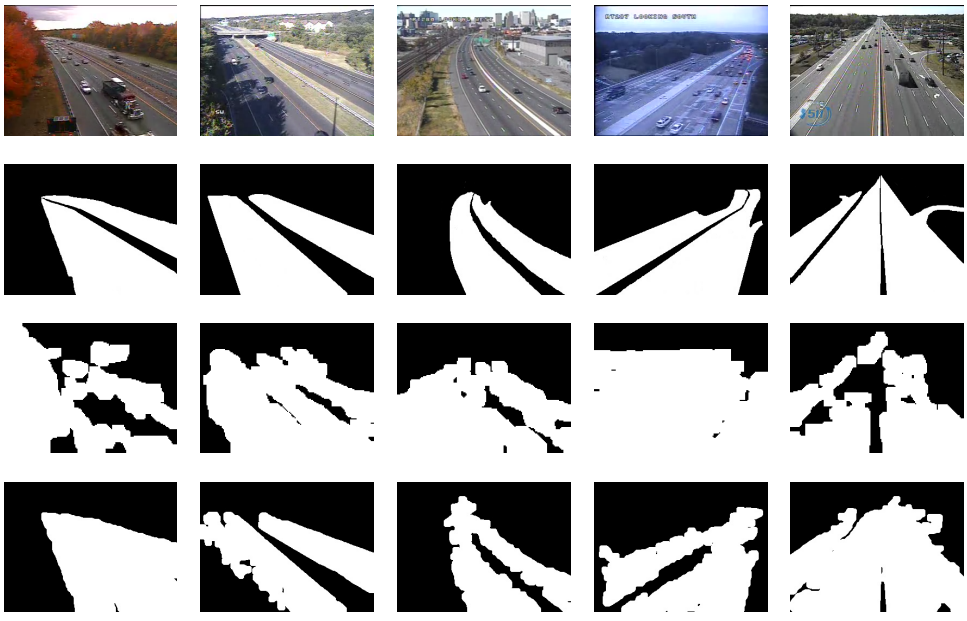
In this subsection, we show some experimental results to evaluate our statistical modeling method for road recognition. The data set we use contains the real traffic surveillance videos from the New Jersey Department of Transportation (NJDOT). To ensure the diversity of the videos, this data set includes ten video sequences with several kinds of resolutions and frame rates, various weather conditions, and different illumination conditions. One frame of each video is displayed in the first and the fifth rows in Figure 5.2. The second and the sixth rows in Figure 5.2 show the ground truth road region masks. The third and the seventh rows in Figure 5.2 present the road recognition result of UFL-HS method [131]. The fourth and the eighth rows in Figure 5.2 show the road region detected by our proposed method. We can see from Figure 5.2, our statistical modeling method can detect the road regions in all the videos accurately.

The feature vector $\vec{x}$ we used in the experiment is the histogram of oriented gradients (HOG) feature [30] calculated from a 4×4 cell surrounding each pixel. The number of the component in the Gaussian mixture model M is 3. The number of frames used for building the model is 50.

We further compare our method with the UFL-HS method [131] quantitatively. The precision, recall, and the F-measure score are popular metrics used to evaluate detection



(a) Video 1 (b) Video 2 (c) Video 3 (d) Video 4 (e) Video 5



(f) Video 6 (g) Video 7 (h) Video 8 (i) Video 9 (j) Video 10

Figure 5.2 The road recognition results. The first and the fifth rows display one video frame from an NJDOT traffic video. The second and the sixth rows show the ground truth road regions. The third and the seventh rows present the road recognition result of UFL-HS method. The fourth and the eighth rows show the road region detected by our proposed method.

Table 5.1 The Quantitative Performance of the Proposed Method

Method	UFL-HS Method			Proposed Method		
Video #	Precision	Recall	F-Score	Precision	Recall	F-Score
1	0.41	0.89	0.56	0.94	0.89	0.91
2	0.73	0.79	0.76	0.96	0.86	0.91
3	0.62	0.68	0.65	0.93	0.78	0.85
4	0.82	0.86	0.84	0.85	0.91	0.88
5	0.68	0.81	0.74	0.88	0.94	0.91
6	0.43	0.62	0.50	0.90	0.98	0.94
7	0.71	0.93	0.80	0.95	0.85	0.90
8	0.53	0.91	0.67	0.91	0.92	0.91
9	0.57	0.95	0.71	0.90	0.88	0.89
10	0.91	0.78	0.84	0.89	0.97	0.92
Average	0.64	0.82	0.72	**0.91**	**0.90**	**0.90**

performance, which are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{5.8}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.9}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.10}$$

where TP, FP, and FN represent the number of true positive, false positive, and false negative detections of the road pixel. In Table. 5.1, we show the quantitative results of our proposed method. We can see the average accuracy of the road region detected by our method is our 90%, which is good enough for it to be used as the RoI.

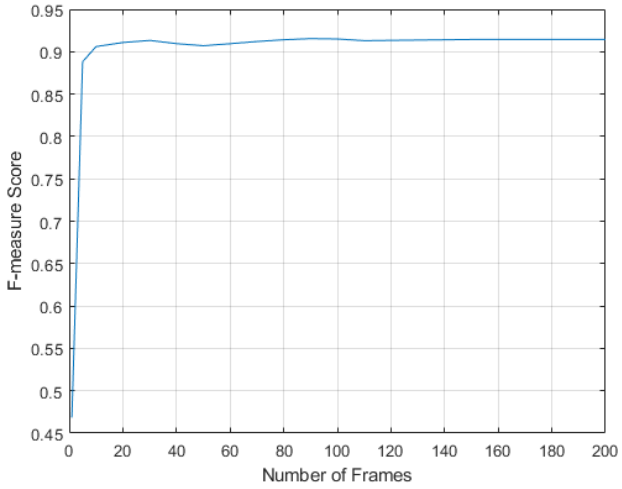


Figure 5.3 The F-measure score of our proposed road recognition method using different number of training frames.

We further investigate the influence of the number of frames used for the model estimation. We change the number of frames used for building the statistical model and calculate the road recognition accuracy. As shown in Figure 5.3, the F-measure score is stable around 0.9 when the number of frames is over 20. Hence, our proposed method does not need to use a large number of frames to build the model, the model estimation process can be fast enough to perform as a pre-processing step of video analysis.

5.4 Conclusions

In this chapter, we have proposed a statistical modeling road recognition method, which switches the info-communications between the RoI selection and video analysis systems from an inter-cognitive communication mode to an intra-cognitive communication mode. The novel road recognition method uses the temporal features in videos instead of manual labeling to generate sample data for statistical modeling and recognizes the road regions which can be used as the RoI in video analysis systems. Our proposed method on one hand improves the road detection accuracy compared to the state-of-the-art method, on the other hand, provides another option for the cognitive info-communications between

the RoI selection and video analysis systems. Our proposed method uses the result of the GFM foreground detection method as guidance to build a statistical road model, and further applies a discriminant function to segment the road regions in the video. Our proposed pruned mixture model is able to correctly segment the shadowed road regions and poor quality roads. The experimental results using the real traffic video sequences from NJDOT verify the robustness and accuracy of our proposed method.

CHAPTER 6

AN INNOVATIVE ANOMALOUS DRIVING DETECTION METHOD IN VIDEO

6.1 Introduction

Traffic video analysis is a compelling topic in computer vision. How to use AI technologies to improve traffic safety is always a major concern of society. Anomalous driving behaviors are one of the most dangerous behaviors in traffic. Every year, thousands of lives are lost in traffic due to anomalous driving. In this chapter, we proposed an anomalous driving detection method, which can detect anomalous driving behaviors in traffic surveillance videos fast and accurately.

Anomaly detection is a very important topic in video analysis. Many methods are proposed to detect the anomalies in videos bases on spatial feature representations [87], [25], [118]. This kind of approach is generally applied to detect all kinds of anomalies without specification. People want to utilize anomaly detection methods in real-world applications. Therefore, more methods are proposed with a specific concentration. To concentrate the attention on anomalous driving behaviors, we proposed our anomalous driving detection method, which integrates the MOT method, the GLV model, and the discriminant function.

We first propose a new multiple object tracking (MOT) method to track the motions of the vehicles in traffic. Many MOT methods are proposed based on a track-by-detection procedure, which needs high computational power. Our proposed MOT method is based on spatial and temporal feature representations, which can process the video frames fast and accurately. By considering the spatial and temporal distance between the objects in two adjacent frames, our MOT method can quickly match the same object in two adjacent frames.

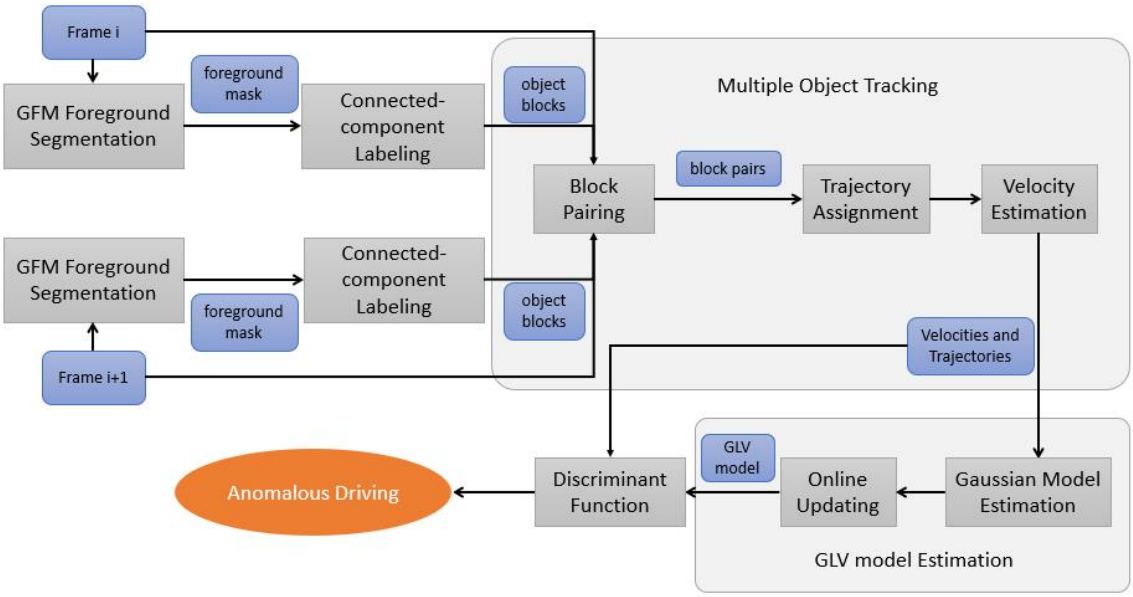


Figure 6.1 The workflow of our proposed anomalous driving detection method using the new MOT method and the novel GLV model.

Second, we use a novel Gaussian local velocity (GLV) modeling method to model the normal driving behaviors in traffic. In order to detect anomalies in traffic, we consider modeling normal behaviors. Everything that cannot be classified in the normal class is an anomaly. As we know, the vehicles in traffic should be driving in a queue. Every vehicle that passes the same location should follow similar speeds and directions, otherwise, traffic accidents may happen. Therefore, we can use a Gaussian distribution to model the normal driving speed and direction at a specific time and location. The Gaussian distribution can be updated online while the speed may change over time.

Third, we use a discriminant function to distinguish the anomalous driving behaviors in videos. With the GLV model, we are able to describe the normal driving behaviors in traffic. Then we can use a discriminant function to classify each moving vehicle in traffic into a normal class or an anomalous class. Figure 6.1 shows the workflow of our proposed anomalous driving detection method.

There are three major contributions of our proposed method. First, the method does not need any training process. The foreground segmentation, MOT process, and GLV modeling method are all unsupervised methods. The models are built when the video is processing. The whole process doe not require any manual label work. This can reduce manual work and increase the general ability of our method. Second, the computational complexity of our proposed method is low and the processing speed is fast. Recently, many anomaly detection methods are proposed based on deep neural networks. However, the implementation of this kind of method requires high-performance GPU, and some of them still cannot process the video in real-time. Unlike deep neural networks, our statistical modeling method has low computational complexity. Even on a normal desktop PC, our method can reach a processing speed of 60 frames per second or faster. Third, our method is robust and generalized. We tested our method on dozens of real traffic videos with different illumination conditions, weather, and resolutions. Our method is able to detect anomalous driving behaviors accurately and robustly. The accuracy is an important criterion to evaluate an anomalous driving detection method. Our method reaches a 100% detection rate without a false alert during testing.

6.2 The New Multiple Object Tracking Method Based on Temporal and Spatial Features

Object tracking is an important topic in video processing. Many methods are proposed to track the moving objects in videos [33], [127], [84]. Multiple object tracking (MOT), which is an extension of single object tracking, plays a more important role in AI applications. People want to catch the trajectories of each individual moving item with MOT, and further tackle some high-level tasks with the tracking results, such as action recognition [26], anomaly detection [43], and object counting [31]. In this subsection, we propose a new MOT method using spatial and temporal information in videos, which can track multiple objects fast and accurately.

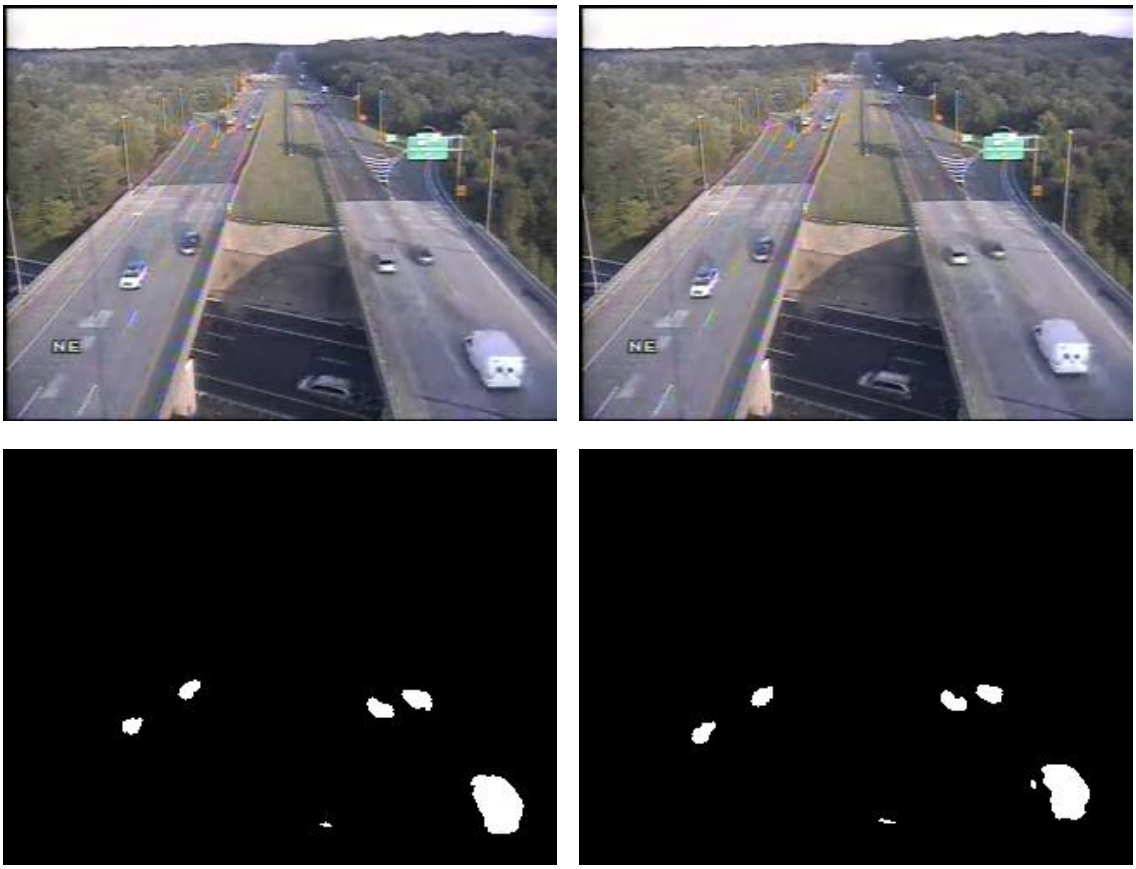


Figure 6.2 Two adjacent video frames from the NJDOT video and their corresponding foreground masks.

To utilize the temporal information in videos, we apply the GFM foreground segmentation method to extract the foreground mask [98], [99]. The GFM method is an unsupervised foreground segmentation method, which can detect the foreground pixels fast and accurately. In the GFM method, foreground pixels are detected using the Bayes decision rule for minimum error, which is described as [5]:

$$C\left(x_{i,j}\right)=\begin{cases}1, & p(\mathbf{x_{i,j}}|\omega_f)P(\omega_f)\geqslant p(\mathbf{x_{i,j}}|\omega_b)P(\omega_b)\\ 0, & Otherwise\end{cases} \tag{6.1}$$

where $x_{i,j}$ represents the pixel at location (i,j), $p(\mathbf{x}|\omega_f)$ and $p(\mathbf{x}|\omega_f)$ means the conditional probability density functions (CPDF) for the foreground and background, respectively,

$P(\omega_f)$ and $P(\omega_b)$ denote the prior probability of the foreground and background, respectively, and C is a binary mask where $C\left(x_{i,j}\right) = 1$ means the pixel at location (i, j) is classified as foreground. The foreground mask represents the moving components in a video. For two adjacent frames $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$, we can get the foreground masks using the GFM method. Figure 6.2 shows two adjacent video frames from an NJDOT traffic video and the corresponding foreground masks detected by the GFM method.

In each foreground mask, there are some foreground pixels, which can indicate the moving objects. A connected-component labeling method [45] is applied to the foreground mask to label every connected region in the foreground mask with a block id. Then we can get two sets of block $S_i = \{B_{i,1}, B_{i,2}, \ldots, B_{i,m}\}$, and $S_{i+1} = \{B_{i+1,1}, B_{i+1,2}, \ldots, B_{i+1,n}\}$ from the two adjacent frames. For every B in set S_i, we pair it with one block B' in set S_{i+1}, which can minimize the distance function $Dist(B, B')$.

$$Dist(B, B') = \frac{Euc\_Dist(B, B')}{Cos\_Dist\left(B, B'\right)} \tag{6.2}$$

$$Euc\_Dist(B, B') = e^{|Cent_B - Cent_{B'}|} \tag{6.3}$$

$$Cos\_Dist(B, B') = \frac{f_B}{||f_B||} \cdot \frac{f_{B'}}{||f_{B'}||} \tag{6.4}$$

where $Cent_B$ and $Cent_{B'}$ mean the centroids of block B and B', f_B and $f_{B'}$ mean the spatial feature vector extracted from the video frame at the corresponding location of block B and B'. In this dissertation, we select the mean and variance of the block area as the spatial features. Note that both the mean and variance are non-negative numbers, the feature vectors are all fall in the same quadrant. Therefore the Cosine similarity is always positive. The novel distance function $Dist(B, B')$ on one hand considers the temporal information, which is the Euclidean distance between the blocks, on the other hand, considers the spatial

information, which is the Cosine similarity between the blocks. After all the blocks in set S_i have paired with a block in set S_{i+1}, we connect the block pairs to trajectories.

Different from the detection based tracking methods, our proposed MOT method does not need to detect each individual item, which can save computational power. Our proposed method is based on a motion-tracking procedure and further involves the spatial features to define an innovative distance function. The combination of the temporal and spatial features can improve the tracking performance with a limited computational requirement.

6.3 The Anomalous Driving Detection using a Novel Gaussian Local Velocity Model

Anomalous driving is one of the most dangerous behaviors in traffic. Many traffic accidents are caused by anomalous driving behaviors, such as wrong-way driving, sudden lane merge, and stopped in traffic. To detect anomalous driving behaviors, we propose the novel Gaussian local velocity (GLV) model to model the normal driving behaviors and use a discriminant function to identify the anomalies in traffic surveillance videos.

Our proposed GLV model uses an unsupervised online updating strategy to establish the model. First, we use the velocities achieved from the MOT method in each frame as input feature vectors to estimate an initial GLV model. The model is updated every frame to satisfy the speed change of traffic. Finally, we apply the discriminant function to the trajectories of each individual foreground object achieved from the MOT method to determine if the driving behavior of that object is anomalous.

Our proposed method is mainly focusing on traffic surveillance video analysis. As we know, vehicles in traffic are moving along the traffic lanes, every vehicle normally moves along a similar direction at the same location. Therefore, we can build a Gaussian distribution for every location in a frame to model the normal velocity. For each location i, j in a frame, the feature vector $\mathbf{x}_{i,j}$ is composed of the magnitude $s_{i,j}$ and the angle $\theta_{i,j}$. The GLV model at location i, j can be described as:

$$V_{i,j} = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}_{i,j} - \mathbf{M}_{i,j})^t \Sigma_{i,j}^{-1}(\mathbf{x}_{i,j} - \mathbf{M}_{i,j})\right\}}{(2\pi)^{d/2} \mid \Sigma_{i,j} \mid^{1/2}} \tag{6.5}$$

where d is the dimensionality of the feature vector $\mathbf{x}_{i,j}$, $\mathbf{M}_{i,j}$ is the mean vector, and $\Sigma_{i,j}$ is the covariance matrix. The GLV model is updated with every moving object passed by the location.

The GLV model describes the normal moving behaviors in traffic. If the motion feature of an object is far from the mean vector of the GLV model, it can be considered as an anomalous motion. We further propose an anomalous driving detection method based on the GLV model.

As we know, the normal driving vehicles should follow a similar direction of the traffic flow. If a vehicle drove in the wrong direction that is away from the traffic flow, it may cause a traffic accident. Our anomalous driving detection is aiming at detecting and alerting this kind of wrong-way driving behavior. We propose a discriminant function to identify the anomalous driving based on the GLV model:

$$D(\mathbf{x}_{i,j}) = \frac{\mathbf{x}_{i,j}}{||\mathbf{x}_{i,j}||} \cdot \frac{\mathbf{M}_{i,j}}{||\mathbf{M}_{i,j}||} \tag{6.6}$$

where $\mathbf{x}_{i,j}$ is the motion feature vector of a vehicle, and $\mathbf{M}_{i,j}$ is the mean vector of the GLV model at location i, j. If the discriminant function $D(\mathbf{x}_{i,j}) < 0$, we classify that vehicle as anomalous driving. If the duration of the anomalous driving behavior of a specific vehicle is longer than the threshold, we identify it as dangerous and send an alert.

6.4 Experiments

To evaluate our proposed anomalous driving detection method, we run experiments on the real traffic video sequences from the New Jersey Department of Transportation (NJDOT). We use a desktop with an Intel Core i7-8700 Processor to implement our proposed method.

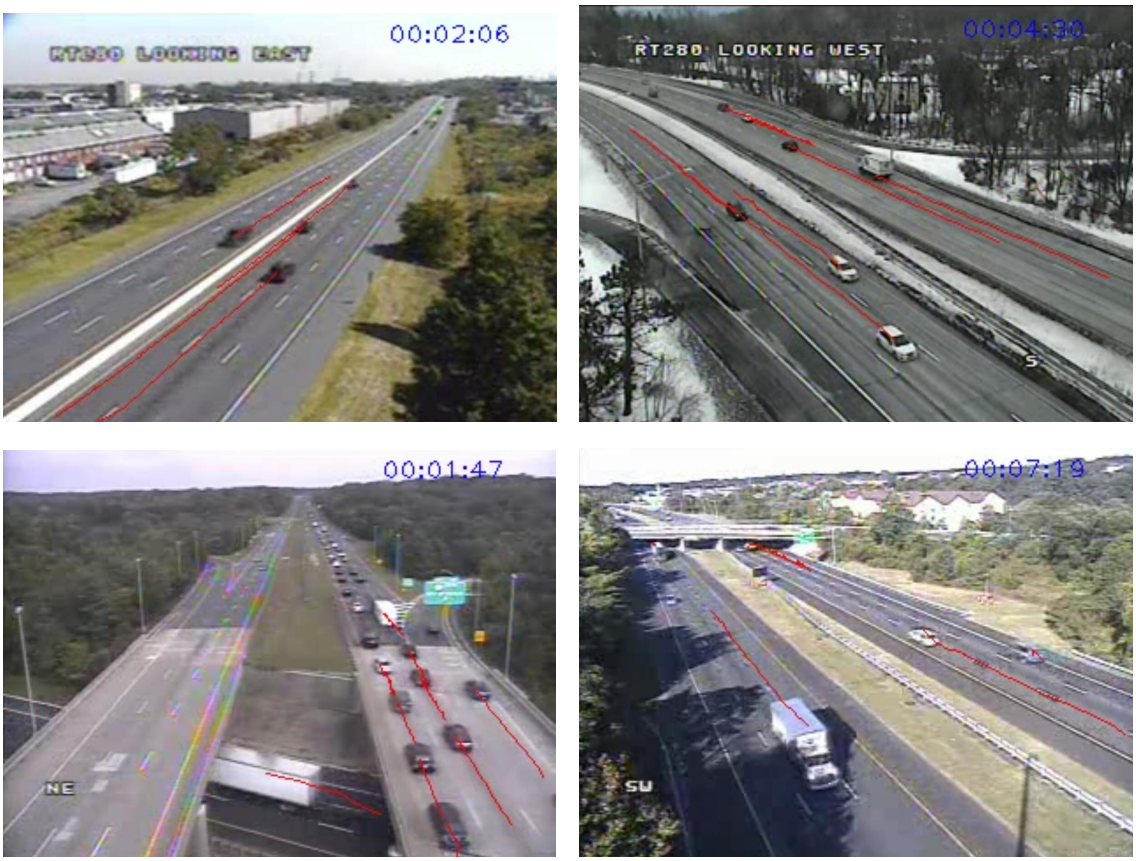


Figure 6.3 The moving trajectories detected by our new MOT method. The red lines indicate the moving trajectories of the vehicles in two seconds.

We first present the results achieved by our proposed MOT method. The processing speed is a very important factor of an MOT method. For a video frame with a resolution of 352×240, the average processing time of our proposed MOT method is 13 ms, which is 77 frames per second or fps. Figure 6.3 shows the vehicle moving trajectories detected by our proposed MOT method. The red lines show 2 seconds of the vehicle moving trajectories. We can see the normal moving vehicles have relatively straight moving trajectories along the road.

The GLV model is built after the moving trajectories are extracted. By utilizing the GLV model, we detect anomalous driving behaviors in traffic videos. Our method can process 65 frames per second (fps) for the videos with the spatial resolution of 352×240,

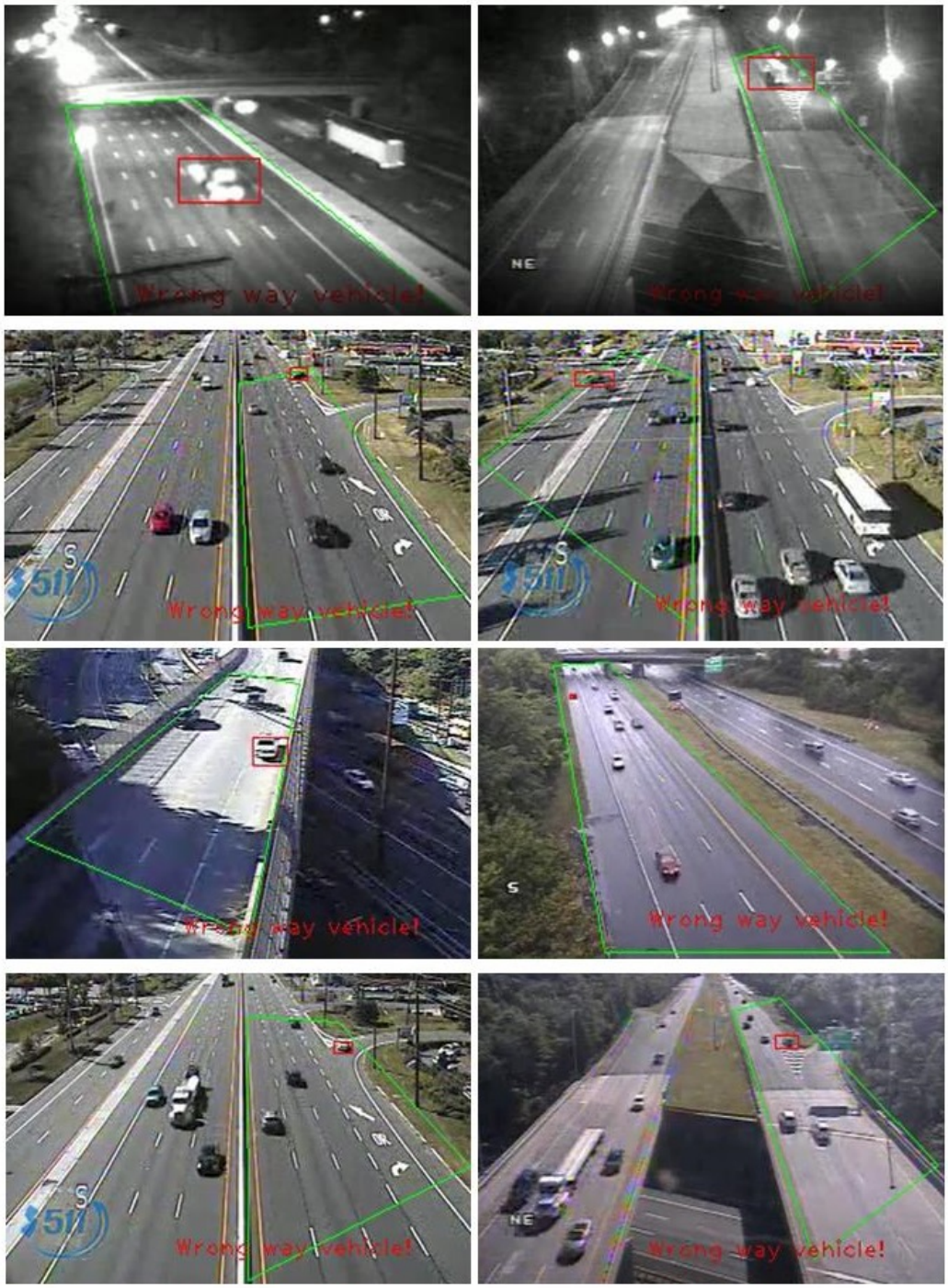


Figure 6.4 The wrong-way driving vehicles detected by our proposed method. The red rectangles are the wrong-way driving vehicles detected by our proposed method, the areas in the green lines are the region of interest (ROI).

which is a general resolution of the traffic surveillance videos. The most commonly seen anomalous driving in traffic is wrong-way driving. Drivers may back up the vehicles in traffic due to missing the existing ramp or entering a wrong ramp. Our proposed method is able to detect these dangerous behaviors in order to minimize traffic accidents happening. We tested our proposed methods on dozens of different real traffic scenarios, our method can detect all the wrong-way driving vehicles without a false alert. Figure 6.4 shows some wrong-way driving vehicles detected by our anomalous driving detection method. The wrong-way driving vehicles are marked with red rectangles. We can see our method is able to deal with both the night videos and daytime videos.

6.5 Conclusion

We proposed a novel anomalous driving behavior detection method for traffic surveillance video analysis in this chapter. The method integrates a new multiple object tracking (MOT) method, a novel Gaussian local velocity (GLV) model, and an anomalous driving discriminant function to detect and alert the anomalous driving behaviors. The MOT method utilizes the spatial and temporal information in video and is able to process the video fast and accurately. The GLV model is built locally using the Gaussian distributions and is updated online. The discriminant function can classify each moving vehicle into a normal driving class or an anomalous driving class. There are three advantages of our proposed method. First, the method proposed is based on statistical modeling. The estimation of the statistical models does not require any labeled data. This can reduce manual labeling work and increase the generalization ability of our method. Second, the computational complexity of our proposed method is low. The anomalous driving detection can process over 60 frames per second on a normal PC. Third, the anomalous driving detection method is accurate and robust to the real-world situation. We tested on dozens of different traffic video scenarios, all the anomalous drivings can be detected without

false alerts. The experimental results using the New Jersey Department of Transportation (NJDOT) real traffic videos show the feasibility of our proposed method.

CHAPTER 7

CONCLUSION AND FUTURE WORK

All the methods proposed in this dissertation are statistical modeling methods for traffic video analysis.

Specifically, Chapter 3 presented a novel foreground detection method for video analysis by integrating color, wavelet, and temporal features. First, a novel Global Foreground Modeling (GFM) method is presented to model the foreground, which estimates a global probability density function for the foreground, and the Bayes decision rule for minimum error is applied to choose one Gaussian density function for a specific time and location. Additionally, a Local Background Modeling (LBM) method is explained by choosing the most significant Gaussian density in the Gaussian mixture model to model the background. Second, the proposed method better satisfies the independence assumption in statistical modeling by applying the unconventional color spaces, such as the YIQ color space, the YCbCr color space, the uncorrelated color space, the independent color space, and the discriminating color space for statistical modeling, in which the covariance matrix is diagonal. Third, to further enhance the discriminatory power of the feature vectors, the horizontal and vertical Haar wavelet features and the temporal difference features are integrated into the color features to build a new 12-dimensional feature vector. The 12-dimensional feature vector thus is able to increase the discriminatory power, which helps our foreground detection method achieve better performance than other popular statistical modeling methods. Finally, the Bayes classifier is applied for the classification of foreground and background pixels. As a result, the proposed method is able to address the challenging problems, such as better satisfying the independence assumption in statistical modeling, insufficient discriminatory power of the input feature vector in the RGB color space, the inappropriate statistical modeling of the foreground, the detection of

the temporary stopped foreground objects, and the final classification of the foreground and background pixels.

In Chapter 4, we proposed a new moving cast shadow detection method using new chromatic criteria and statistical modeling to enhance the foreground detection results. First, a set of new chromatic criteria is proposed for candidate shadow pixel detection in the HSV color space. The new chromatic criteria are more suitable for videos of different quality. Second, a new shadow region detection method is presented to detect continuous candidate shadow regions. The shadow region detection method can solve the problem that the shadow edge is easy to be missed in the pixel-based detection methods. Third, a statistical shadow model is built using a single Gaussian distribution for shadow pixel classification. Fourth, an aggregated shadow detection method is presented to generate the foreground without shadows. By aggregating the shadow detection results of the previous three methods, we can obtain a more accurate shadow detection result. With the shadow detection method, we can remove the cast shadows which may influence the video analysis performance.

Chapter 5 represented a road recognition method using statistical modeling. The road regions are recognized as the Region of Interest (RoI) for traffic video analysis systems. Our proposed road recognition method has three major contributions. First, we introduce a temporal feature guided statistical modeling method. The method uses temporal features in videos as guidance to automatically extract data to build the statistical model. Second, we propose a road recognition method, which can detect the road regions in videos. A discriminant function is applied to segment the road regions based on their feature representations. Third, the automatic RoI detection procedure switches the info-communications between the RoI selection and video analysis systems from an inter-cognitive communication mode to an intra-cognitive communication mode. This largely reduces the human work and subjective error in the traffic video analysis systems.

Chapter 6 introduced a novel anomalous driving detection method in videos, which can detect unsafe anomalous driving behaviors using statistical models. First, a new Multiple Object Tracking (MOT) method is proposed. The new MOT method can detect and track the moving foreground objects with low computational complexity. Second, a novel Gaussian Local Velocity (GLV) modeling method is presented. The GLV model is used to model the normal driving behaviors in traffic. The advantage of such a model is that the estimation of the model does not require any labeled data. Third, a discriminant function is proposed to detect anomalous driving behaviors. By applying the discriminant function, our novel anomalous driving detection method is able to detect the anomalies in traffic surveillance videos fast and accurately.

The experimental results using some publicly available videos and the NJDOT videos have shown that our proposed methods can achieve better performance than the other popular methods. Our future work will focus on developing more statistical modeling methods for video analysis applications, and integrate our proposed methods into real world applications.

REFERENCES

[1] Norel Ya Qine Abderrahim, Saadane Abderrahim, and Azmi Rida. Road segmentation using u-net architecture. In *IEEE International conference of Moroccan Geomatics (Morgeo)*, May 11-13, 2020, Casablanca, Morocco.

[2] José M Álvarez Alvarez and Antonio M Lopez. Road detection based on illuminant invariance. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):184–193, 2010.

[3] Mohamed Aly. Real time detection of lane markers in urban streets. In *IEEE Intelligent Vehicles Symposium (IV)*, June 4-6, 2008, Eindhoven, Netherlands.

[4] Ariel Amato, Mikhail G Mozerov, Andrew D Bagdanov, and Jordi Gonzalez. Accurate moving cast shadow suppression based on local color constancy detection. *IEEE Transactions on Image Processing*, 20(10):2954–2966, 2011.

[5] Keith D. Copsey Andrew R. Webb. *Statistical Pattern Recognition, 3rd Edition*. New York, NY, USA: John Wiley & Sons, 2011.

[6] Sugata Banerji, Atreyee Sinha, and Chengjun Liu. New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. *Neurocomputing*, 117:173–185, 2013.

[7] Péter Baranyi and Ádám Csapó. Definition and synergies of cognitive infocommunications. *Acta Polytechnica Hungarica*, 9(1):67–83, 2012.

[8] Péter Baranyi, Adam Csapo, and Gyula Sallai. *Cognitive Infocommunications*. Berlin, Germany: Springer, 2015.

[9] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6):1709–1724, 2011.

[10] Kuldeep Marotirao Biradar, Ayushi Gupta, Murari Mandal, and Santosh Kumar Vipparthi. Challenges in time-stamp aware anomaly detection in traffic videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, June 16-20, 2019, Long Beach, CA, USA.

[11] Aissa Boulmerka and Mohand Saïd Allili. Foreground segmentation in videos combining general gaussian mixture modeling and spatial information. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6):1330–1345, 2018.

[12] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent Patents on Computer Science*, 1(3):219–237, 2008.

[13] Thierry Bouwmans, Caroline Silva, Cristina Marghes, Mohammed Sami Zitouni, Harish Bhaskar, and Carl Frelicot. On the role and the importance of features for background modeling and foreground detection. *Computer Science Review*, 28:26–91, 2018.

[14] Thierry Bouwmans, Andrews Sobral, Sajid Javed, Soon Ki Jung, and El-Hadi Zahzah. Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset. *Computer Science Review*, 23:1–71, 2017.

[15] Guillem Braso and Laura Leal-Taixe. Learning a neural solver for multiple object tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 13-19, 2020, Seattle, WA, USA.

[16] Ron Brinkmann. *The art and science of digital compositing: Techniques for visual effects, animation and motion graphics*. Burlington, MA, USA: Morgan Kaufmann, 2008.

[17] Sebastian Brutzer, Benjamin Höferlin, and Gunther Heidemann. Evaluation of background subtraction techniques for video surveillance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-25, 2011, Colorado Springs, CO, USA.

[18] Elad Bullkich, Idan Ilan, Yair Moshe, Yacov Hel-Or, and Hagit Hel-Or. Moving shadow detection by nonlinear tone-mapping. In *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, April 11-13, 2012, Vienna, Austria.

[19] Luca Caltagirone, Mauro Bellone, Lennart Svensson, and Mattias Wahde. Lidar-camera fusion for road detection using fully convolutional neural networks. *Robotics and Autonomous Systems*, 111:125–131, 2019.

[20] Chia-Chih Chen and Jake K Aggarwal. Human shadow removal with unknown light source. In *20th International Conference on Pattern Recognition (ICPR)*, August 23-26, 2010, Istanbul, Turkey.

[21] Chun-Ting Chen, Chung-Yen Su, and Wen-Chung Kao. An enhanced segmentation on vision-based shadow removal for vehicle detection. In *International Conference on Green Circuits and Systems (ICGCS)*, June 21-23, 2010, Shanghai, China.

[22] Shuo Chen and Chengjun Liu. Clustering-based discriminant analysis for eye detection. *IEEE Transactions on Image Processing*, 23(4):1629–1638, 2014.

[23] Shuo Chen and Chengjun Liu. Eye detection using discriminatory haar features and a new efficient SVM. *Image and Vision Computing*, 33:68–77, 2015.

[24] Zhe Chen, Jing Zhang, and Dacheng Tao. Progressive lidar adaptation for road detection. *IEEE/CAA Journal of Automatica Sinica*, 6(3):693–702, 2019.

[25] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. *IEEE Transactions on Image Processing*, 24(12):5288–5301, 2015.

[26] Jin Choi, Yong-il Cho, Kyusung Cho, Sujung Bae, and Hyun Seung Yang. A view-based multiple objects tracking and human action recognition for interactive virtual environments. *International Journal of Virtual Reality*, 7(3):71–76, 2008.

[27] Qi Chu, Wanli Ouyang, Hongsheng Li, Xiaogang Wang, Bin Liu, and Nenghai Yu. Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism. In *IEEE International Conference on Computer Vision (ICCV)*, October 22-29, 2017, Venice, Italy.

[28] Pierre Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, 1994.

[29] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, 2003.

[30] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-26, 2005, San Diego, CA, USA.

[31] Carlos R. del-Blanco, Fernando Jaureguizar, and Narciso García. An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications. *IEEE Transactions on Consumer Electronics*, 58(3):857–862, 2012.

[32] Keval Doshi and Yasin Yilmaz. Fast unsupervised anomaly detection in traffic videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, June 13-19, 2020, Seattle, WA, USA.

[33] Litong Fan, Zhongli Wang, Baigen Cai, Chuanqi Tao, Zhiyi Zhang, Yinling Wang, Shanwen Li, Fengtian Huang, Shuangfu Fu, and Feng Zhang. A survey on multiple object tracking algorithm. In *IEEE International Conference on Information and Automation (ICIA)*, August 1-3, 2016, Ningbo, China.

[34] Liuzhi Fang, Wangyun Qiong, and Youzhi Sheng. A method to segment moving vehicle cast shadow based on wavelet transform. *Pattern Recognition Letters*, 29(16):2182–2188, 2008.

[35] David A. Forsyth and Jean Ponce. *Computer Vision - A Modern Approach, Second Edition*. London, UK: Pitman, 2012.

[36] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Cambridge, MA, USA: Academic Press, 2013.

[37] Syed Asif Mehmood. Gilani, Ioannis Kostopoulos, and Athanassios Skodras. Color image-adaptive watermarking. In *14th International Conference on Digital Signal Processing (DSP)*, July 1-3, 2002, Santorini, Greece.

[38] Vitor Gomes, Pablo Barcellos, and Jacob Scharcanski. Stochastic shadow detection using a hypergraph partitioning approach. *Pattern Recognition*, 63:30–44, 2017.

[39] Rafael C. González and Richard E. Woods. *Digital Image Processing, 4th Edition*. London, UK: Pearson Education, 2018.

[40] Nil Goyette, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, and Prakash Ishwar. Changedetection. net: A new change detection benchmark dataset. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, June 16-21, 2012, Providence, RI, USA.

[41] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2956–2967, 2013.

[42] Siqiu Guo, Tao Zhang, YuLong Song, and Feng Qian. Color feature-based object tracking through particle swarm optimization with improved inertia weight. *Sensors*, 18(4):1292, 2018.

[43] Mei Han and Yihong Gong. Real-time multiple-object tracking and anomaly detection. In Rainer Lienhart, Noboru Babaguchi, and Edward Y. Chang, editors, *Storage and Retrieval Methods and Applications for Multimedia (SRMAM)*, January 18, 2005, San Jose, CA, USA.

[44] Eric Hayman and Jan-Olof Eklundh. Statistical background subtraction for a mobile observer. In *9th IEEE International Conference on Computer Vision (ICCV)*, October 14-17, 2003, Nice, France.

[45] Lifeng He, Yuyan Chao, Kenji Suzuki, and Kesheng Wu. Fast connected-component labeling. *Pattern Recognition*, 42(9):1977–1987, 2009.

[46] Mohamed A Helala, Ken Q Pu, and Faisal Z Qureshi. Road boundary detection in challenging scenarios. In *9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, September 18-21, 2012, Beijing, China.

[47] Aharon Bar Hillel, Ronen Lerner, Dan Levi, and Guy Raz. Recent progress in road and lane detection: a survey. *Machine Vision and Applications*, 25(3):727–745, 2014.

[48] Jun-Wei Hsieh, Wen-Fong Hu, Chia-Jung Chang, and Yung-Sheng Chen. Shadow elimination for effective moving object detection by gaussian shadow modeling. *Image and Vision Computing*, 21(6):505–516, 2003.

[49] Jia-Bin Huang and Chu-Song Chen. Moving cast shadow detection using physics-based features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-25, 2009, Miami, Florida, USA.

[50] Shih-Shinh Huang, Li-Chen Fu, and Pei-Yung Hsiao. Region-level motion-based background modeling and subtraction using mrfs. *IEEE Transactions on Image Processing*, 16(5):1446–1456, 2007.

[51] Sahin Isik, Kemal Özkan, Serkan Günal, and Ömer Nezih Gerek. Swcd: a sliding window and self-regulated learning-based background updating method for change detection in videos. *Journal of Electronic Imaging*, 27(2):023002, 2018.

[52] Wei Ji and Yong Zhao. Moving cast shadow detection using joint color and texture features based on direction and distance. In *2nd IEEE International Conference on Computer and Communications (ICCC)*, October 14-17, 2016, Chengdu, China.

[53] Felipe Jorquera, Sergio Hernández, and Diego Vergara. Probability hypothesis density filter using determinantal point processes for multi object tracking. *Computer Vision and Image Understanding*, 183:33–41, 2019.

[54] Juha Karhunen, Erkki Oja, Liuyue Wang, Ricardo Vigário, and Jyrki Joutsensalo. A class of neural networks for independent component analysis. *IEEE Transactions on Neural Networks*, 8(3):486–504, 1997.

[55] Salman H Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3:431–446, 2016.

[56] Du Yong Kim, Ba-Ngu Vo, Ba-Tuong Vo, and Moongu Jeon. A labeled random finite set online multi-object tracker for video data. *Pattern Recognition*, 90:377–389, 2019.

[57] Hui Kong, Jean-Yves Audibert, and Jean Ponce. Vanishing point detection for road detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-25, 2009, Miami, Florida, USA.

[58] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *11th European Conference on Computer Vision (ECCV)*, September 5-11, 2010, Heraklion, Crete, Greece.

[59] Hasup Lee, HyungSeok Kim, and Jee-In Kim. Background subtraction using background sets with image-and color-space reduction. *IEEE Transactions on Multimedia*, 18(10):2093–2103, 2016.

[60] Jae-Seol Lee and Tae-Hyoung Park. Fast road detection by cnn-based camera-lidar fusion and spherical coordinate transformation. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–9, 2020.

[61] Woochul Lee and Bin Ran. Bidirectional roadway detection for traffic surveillance using online cctv videos. In *IEEE Intelligent Transportation Systems Conference (ITSC)*, September 17-20, 2006, Toronto, Ontario, Canada.

[62] Alessandro Leone and Cosimo Distante. Shadow detection for moving objects based on texture analysis. *Pattern Recognition*, 40(4):1222–1233, 2007.

[63] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu, and Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, 2004.

[64] Yong Li, Weili Ding, XuGuang Zhang, and Zhaojie Ju. Road detection algorithm for autonomous navigation systems based on dark channel prior and vanishing point in complex road scenes. *Robotics and Autonomous Systems*, 85:1–11, 2016.

[65] Yong Li, Guofeng Tong, Anan Sun, and Weili Ding. Road extraction algorithm based on intrinsic image and vanishing point for unstructured road image. *Robotics and Autonomous Systems*, 109:86–96, 2018.

[66] Chengjun Liu. A bayesian discriminating features method for face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.

[67] Chengjun Liu. Learning the uncorrelated, independent, and discriminating color spaces for face recognition. *IEEE Transactions on Information Forensics and Security*, 3(2):213–222, 2008.

[68] Chengjun Liu. *Recent Advances in Intelligent Image Search and Video Retrieval*. Berlin, Germany: Springer, 2017.

[69] Chengjun Liu and Jian Yang. ICA color space for pattern recognition. *IEEE Transactions on Neural Networks*, 20(2):248–257, 2009.

[70] Mingzhou Liu, Qiannan Jiang, and Jing Hu. Detection of highway lane lines and drivable regions based on dynamic image enhancement algorithm under unfavorable vision. *Computers and Electrical Engineering*, 89:106911, 2021.

[71] Qingfeng Liu and Chengjun Liu. A novel locally linear knn method with applications to visual recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 28(9):2010–2021, 2017.

[72] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection – a new baseline. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 18-22, 2018, Salt Lake City, UT, USA.

[73] Ezequiel López-Rubio, Miguel A Molina-Cabello, Rafael Marcos Luque-Baena, and Enrique Domínguez. Foreground detection by competitive learning for varying input distributions. *International Journal of Neural Systems*, 28(05):1750056, 2018.

[74] Yecheng Lyu, Lin Bai, and Xinming Huang. Road segmentation using cnn and distributed lstm. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 26-29, 2019, Sapporo, Japan.

[75] Rohini Mahajan and Abhijeet Bajpayee. A survey on shadow detection and removal based on single light source. In *9th IEEE International Conference on Intelligent Systems and Control (ISCO)*, January 9-10, 2015, Coimbatore, India.

[76] Andrii Maksai, Xinchao Wang, Francois Fleuret, and Pascal Fua. Non-markovian globally consistent multi-object tracking. In *IEEE International Conference on Computer Vision (ICCV)*, October 22-29, 2017, Venice, Italy.

[77] Isabel Martins, Pedro Carvalho, Luís Corte-Real, and José Luis Alba-Castro. Bmog: boosted gaussian mixture model with controlled complexity. In *8th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, June 20-23, 2017, Faro, Portugal.

[78] José Melo, Andrew Naftel, Alexandre Bernardino, and José Santos-Victor. Detection and classification of highway lanes using vehicle motion trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 7(2):188–200, 2006.

[79] Vipul H Mistry and Ramji Makwana. Survey: Vision based road detection techniques. *International Journal of Computer Science and Information Technologies*, 5:4741–4747, 2014.

[80] Anurag Mittal and Daniel P. Huttenlocher. Scene modeling for wide area surveillance and image synthesis. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, June 13-15, 2000, Hilton Head, SC, USA.

[81] Sohail Nadimi and Bir Bhanu. Physical models for moving shadow and object detection in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1079–1087, 2004.

[82] Aparajit Narayan, Elio Tuci, Frederic Labrosse, and Muhanad H Mohammed Alkilabi. Road detection using convolutional neural networks. In *14th European Conference Artificial Life (ECAL)*, September 4-8, 2017, Lyon, France.

[83] Trong-Nguyen Nguyen and Jean Meunier. Anomaly detection in video sequence with appearance-motion correspondence. In *IEEE International Conference on Computer Vision (ICCV)*, October 27-November 2, 2019, Seoul, Korea (South).

[84] Hui-Lee Ooi, Guillaume-Alexandre Bilodeau, and Nicolas Saunier. Supervised and unsupervised detections for multiple object tracking in traffic scenes: A comparative study. In Aurélio Campilho, Fakhri Karray, and Zhou Wang, editors, *17th International Conference on Image Analysis and Recognition (ICIAR)*, June 24-26, 2020, Póvoa de Varzim, Portugal.

[85] Dominika Palivcova, Miroslav Macik, and Zdenek Mikovec. Susy: Surveillance system for hospitals. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, September 11-14, 2017, Debrecen, Hungary.

[86] Megha Pandey and Svetlana Lazebnik. Scene recognition and weakly supervised object localization with deformable part-based models. In *IEEE International Conference on Computer Vision, (ICCV)*, November 6-13, 2011, Barcelona, Spain.

[87] Neha Patil and Prabir Kumar Biswas. A survey of video datasets for anomaly detection in automated surveillance. In *6th International Symposium on Embedded Computing and System Design (ISED)*, December 15-17, 2016, Patna, India.

[88] Jau-Woei Perng, Ya-Wen Hsu, Ya Zhu Yang, Chia-Yen Chen, and Tang-Kai Ying. Development of an embedded road boundary detection system based on deep learning. *Image and Vision Computing*, 100:103935, 2020.

[89] Andrea Prati, Ivana Mikic, Mohan M Trivedi, and Rita Cucchiara. Detecting moving shadows: algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):918–923, 2003.

[90] Ajit Puthenputhussery, Qingfeng Liu, and Chengjun Liu. A sparse representation model using the complete marginal fisher analysis framework and its applications to visual recognition. *IEEE Transactions on Multimedia*, 19(8):1757–1770, 2017.

[91] Ming Qin, Yao Lu, Huijun Di, and Wei Huang. A background basis selection-based foreground detection method. *IEEE Transactions on Multimedia*, 18(7):1283–1296, 2016.

[92] Liangliang Ren, Jiwen Lu, Zifeng Wang, Qi Tian, and Jie Zhou. Collaborative deep reinforcement learning for multi-object tracking. In *15th European Conference on Computer Vision (ECCV)*, September 8-14, 2018, Munich, Germany.

[93] Elena Salvador, Andrea Cavallaro, and Touradj Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, 2004.

[94] Andres Sanin, Conrad Sanderson, and Brian C Lovell. Shadow detection: A survey and comparative evaluation of recent methods. *Pattern Recognition*, 45(4):1684–1695, 2012.

[95] Andres Sanin, Conrad Sanderson, and Brian C Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. In *20th International Conference on Pattern Recognition (ICPR)*, August 23-26, 2010, Istanbul, Turkey.

[96] Young-Woo Seo and Ragunathan Raj Rajkumar. Detection and tracking of boundary of unmarked roads. In *17th International Conference on Information Fusion (FUSION)*, July 7-10, 2014, Salamanca, Spain.

[97] Yaser Sheikh and Mubarak Shah. Bayesian object detection in dynamic scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-26, 2005, San Diego, CA, USA.

[98] Hang Shi and Chengjun Liu. A new foreground segmentation method for video analysis in different color spaces. In *24th International Conference on Pattern Recognition (ICPR)*, August 20-24, 2018, Beijing, China.

[99] Hang Shi and Chengjun Liu. A new global foreground modeling and local background modeling method for video analysis. In *14th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM)*, July 15-19, 2018, New York, NY, USA.

[100] Peichung Shih and Chengjun Liu. Comparative assessment of content-based face image retrieval in different color spaces. *International Journal of Pattern Recognition and Artificial Intelligence*, 19(7):873–893, 2005.

[101] Peichung Shih and Chengjun Liu. Face detection using discriminating feature analysis and support vector machine. *Pattern Recognition*, 39(2):260–276, 2006.

[102] Atreyee Sinha, Sugata Banerji, and Chengjun Liu. New color GPHOG descriptors for object and scene image classification. *Machine Vision and Applications*, 25(2):361–375, 2014.

[103] Andrews Sobral and Antoine Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4–21, 2014.

[104] Jongin Son, Hunjae Yoo, Sanghoon Kim, and Kwanghoon Sohn. Real-time illumination invariant lane detection for lane departure warning system. *Expert Systems with Applications*, 42(4):1816–1824, 2015.

[105] Miguel Angel Sotelo, Francisco Javier Rodriguez, and Luis Magdalena. Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 5(2):69–83, 2004.

[106] Pierre-Luc St-Charles, Guillaume-Alexandre Bilodeau, and Robert Bergevin. Universal background subtraction using word consensus models. *IEEE Transactions on Image Processing*, 25(10):4768–4781, 2016.

[107] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, June 23-25, 1999, Ft. Collins, CO, USA.

[108] Bangyu Sun and Shutao Li. Moving cast shadow detection of vehicle using combined color models. In *Chinese Conference on Pattern Recognition (CCPR)*, October 21-23, 2010, Chongqing, China.

[109] Ceryen Tan, Tsai Hong, Tommy Chang, and Michael Shneier. Color model-based real-time learning for road following. In *IEEE Intelligent Transportation Systems Conference (ITSC)*, September 17-20, 2006, Toronto, Ontario, Canada.

[110] Siyu Tang, Bjoern Andres, Miykhaylo Andriluka, and Bernt Schiele. Subgraph decomposition for multi-target tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015, Boston, MA, USA.

[111] Yao Tang, Lin Zhao, Shanshan Zhang, Chen Gong, Guangyu Li, and Jian Yang. Integrating prediction and reconstruction for anomaly detection. *Pattern Recognition Letters*, 129:123–130, 2020.

[112] Guofeng Tong, Yong Li, Anan Sun, and Yuebin Wang. Shadow effect weakening based on intrinsic image extraction with effective projection of logarithmic domain for road scene. *Signal, Image and Video Processing*, 14(4):1–9, 2019.

[113] Sriram Varadarajan, Paul C. Miller, and Huiyu Zhou. Region-based mixture of gaussians modelling for foreground detection in dynamic scenes. *Pattern Recognition*, 48(11):3488–3503, 2015.

[114] Arun Varghese and G Sreelekha. Sample-based integrated background subtraction and shadow detection. *IPSJ Transactions on Computer Vision and Applications*, 9(1):25, 2017.

[115] Tomás F Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):682–695, 2018.

[116] Yago Vicente, F Tomas, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection. In *IEEE International Conference on Computer Vision (ICCV)*, December 7-13, 2015, Santiago, Chile.

[117] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, December 8-14, 2001, Kauai, HI, USA.

[118] Hung Vu, Tu Dinh Nguyen, Trung Le, Wei Luo, and Dinh Q. Phung. Robust anomaly detection in videos using multilevel representations. In *33rd AAAI Conference on Artificial Intelligence (AAAI)*, January 27-February 1, 2019, Honolulu, Hawaii, USA.

[119] Ende Wang, Yong Li, Anan Sun, Huashuai Gao, Jingchao Yang, and Zheng Fang. Road detection based on illuminant invariance and quadratic estimation. *Optik*, 185:672–684, 2019.

[120] Ende Wang, Anan Sun, Yong Li, Xukui Hou, and Yalong Zhu. Fast vanishing point detection method based on road border region estimation. *IET Image Processing*, 12(3):361–373, 2017.

[121] Yang Wang. Real-time moving vehicle detection with cast shadow removal in video based on conditional random field. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(3):437–441, 2009.

[122] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. Cdnet 2014: an expanded change detection benchmark dataset. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, June 23-28, 2014, Columbus, OH, USA.

[123] Yue Wang, Eam Khwang Teoh, and Dinggang Shen. Lane detection and tracking using b-snake. *Image and Vision Computing*, 22(4):269–280, 2004.

[124] Zheng Wang, Ming-Jun Lai, Zhaosong Lu, Wei Fan, Hasan Davulcu, and Jieping Ye. Orthogonal rank-one matrix pursuit for low rank matrix completion. *SIAM Journal on Scientific Computing*, 37(1), 2015.

[125] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

[126] Dong Xu, Jianzhuang Liu, Xuelong Li, Zhengkai Liu, and Xiaoou Tang. Insignificant shadow detection for video segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(8):1058–1064, 2005.

[127] Yingkun Xu, Xiaolong Zhou, Shengyong Chen, and Fenfen Li. Deep learning for multiple object tracking: a survey. *IET Computer Vision*, 13(4):355–368, 2019.

[128] Yong Xu, Jixiang Dong, Bob Zhang, and Daoyun Xu. Background modeling methods in video analysis: A review and comparative evaluation. *CAAI Transactions on Intelligence Technology*, 1(1):43–60, 2016.

[129] Yongzheng Xu, Guizhen Yu, Xinkai Wu, Yunpeng Wang, and Yalong Ma. An enhanced viola-jones vehicle detection method from unmanned aerial vehicles imagery. *IEEE Transactions on Intelligent Transportation System*, 18(7):1845–1856, 2017.

[130] Jian Yang, Chengjun Liu, and Lei Zhang. Color space normalization: Enhancing the discriminating power of color spaces for face recognition. *Pattern Recognition*, 43(4):1454–1466, 2010.

[131] Junkang Zhang, Siyu Xia, Kaiyue Lu, Hong Pan, and A Kai Qin. Robust road detection from a single image. In *23rd International Conference on Pattern Recognition (ICPR)*, December 4-8, 2016, Cancún, Mexico.

[132] Pengxiang Zhao, Yanyun Zhao, and Anni Cai. Hierarchical codebook background model using haar-like features. In *3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, September 21-23, 2012, Beijing, China.

[133] Shengyan Zhou, Jianwei Gong, Guangming Xiong, Huiyan Chen, and Karl Iagnemma. Road detection using support vector machine based on online learning and evaluation. In *IEEE Intelligent Vehicles Symposium (IV)*, June 21-24, 2010, La Jolla, CA, USA.

[134] Shengyan Zhou, Yanhua Jiang, Junqiang Xi, Jianwei Gong, Guangming Xiong, and Huiyan Chen. A novel lane detection based on geometrical model and gabor filter. In *IEEE Intelligent Vehicles Symposium (IV)*, June 21-24, 2010, La Jolla, CA, USA.

[135] Yong Zhou, Rong Xu, Xiaofeng Hu, and Qingtai Ye. A robust lane detection and tracking method based on computer vision. *Measurement Science and Technology*, 17(4):736, 2006.

[136] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *17th International Conference on Pattern Recognition (ICPR)*, August 23-26, 2004, Cambridge, UK.

[137] Zoran Zivkovic and Ferdinand van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2006.